Quaternary Science Reviews 212 (2019) 1–17

Contents lists available at ScienceDirect

Quaternary Science Reviews

journal homepage: www.elsevier.com/locate/quascirev

Reconstructing past biomes states using machine learning and modern pollen assemblages: A case study from Southern Africa

Magdalena K. Sobol [a, *], Louis Scott [b], Sarah A. Finkelstein [a]

[a] Department of Earth Sciences, University of Toronto, 22 Russell St, Toronto, M5S 3B1, Canada
[b] Department of Plant Sciences, University of the Free State, PO Box 339, Bloemfontein, 9300, South Africa

ARTICLE INFO

Article history:
Received 27 September 2018
Received in revised form
28 February 2019
Accepted 25 March 2019
Available online 4 April 2019

Keywords:
Pollen datasets
Data analysis
Objective classification
Biomes
Vegetation dynamics
Vegetation reconstructions
Late Pleistocene
Holocene

ABSTRACT

Fossil pollen assemblages can assist in understanding biome responses to global climate change if there is reasonable probability that they represent specific biomes or bioregions. In this paper, we introduce a novel probabilistic presentation of pollen data and biome assignment. We apply a recently developed pollen-based vegetation classification method utilizing supervised machine learning to Southern Africa modern pollen assemblages. We present an updated modern pollen dataset from Southern Africa, linking the sites to previously defined vegetation units and, ultimately, we generate probabilistic classification for fossil assemblages to reconstruct past vegetation.

The modern pollen dataset (N = 211 sites) represents a long vegetation gradient, from desert to forest biomes, capturing broad climate gradients ranging from arid to subtropical. We validate two models using Random Forest algorithm to classify modern vegetation at different spatial resolutions: subcontinental (biomes) and regional (bioregions). When the modern pollen assemblages (N = 164 sites) are used to predict the vegetation types, the classification models are correct in a number of cases. In our dataset of 164 sites, the classification model correctly classifies pollen assemblages from savanna (91% correct), grassland (87%), and coastal forest (82%) vegetation types, while the best results for classification of regional vegetation are achieved for sub-humid savanna (95%), dry savanna (95%), coastal forest (91%), and wet grassland (90%).

We apply the models to a fossil pollen sequence at Wonderkrater in the South African savanna, to reconstruct subcontinental and regional changes in past vegetation states over the last 60 000 years. The most probable vegetation state dominating the region since the Late Pleistocene is sub-humid savanna yet grassland occurred at times associated with high vegetation variability. Within the record, the most frequent and amplified variability in the inferred vegetation states occurred during the transitional phase between the Late Pleistocene and the Holocene. The machine learning approach for reconstructing past vegetation, offers a more complex and nuanced view of past vegetation dynamics and has the potential to support quantitative proxy-based techniques for palaeoclimatic reconstructions.

1. Introduction

Modeling complex biomes using multivariate proxy data is ecologically challenging and computationally expensive. At the time of its development, the biomization method (Prentice et al., 1992, 1996) was a revolutionary approach to modeling biomes from pollen data. The method rests on the assumption that the functional relationship between form and function of few key plants may be substituted for biomes and biome modeling. Biomes are classified using pollen assemblages through plant functional types (PFTs); to link pollen assemblages to biomes, two binary matrices assigning pollen assemblages to PFTs, and PTFs to biomes are multiplied (Prentice et al., 1992). Thus, the biomization method reduces large complex datasets to a smaller number of representative plants.

The method, however, relies on few key pollen taxa to represent biomes. Methods considering whole pollen datasets may provide additional nuances particularly applicable to periods of high climatic variability (Williams et al., 2004). Moreover, PFTs created for one region cannot be easily applied to another region. Thus, new PFTs must be created for new contexts. Methodological biases can

* Corresponding author.
E-mail address: magdalena.sobol@mail.utoronto.ca (M.K. Sobol).

https://doi.org/10.1016/j.quascirev.2019.03.027
0277-3791/

further limit the biomization method. Firstly, the method involves a degree of subjectivity as it relies on expert knowledge; data matrices linking pollen to PTFs, and PFTs to biomes are hand-defined by an expert leaving room for alternative interpretations. Furthermore, in some cases of biomization approaches, taxa are selectively removed from the datasets, which may introduce another component of subjectivity. Secondly, the biomization method does not split data into clear training and test sets (Prentice et al., 1992), meaning that there is limited opportunity for independent validation of the model's performance or properties. Thirdly, biomization relies on hand-tuning of models which may lead to models that are fit too closely to a limited set of data.

Moreover, models created via the biomization method are typically assessed in qualitative terms. Definition of precision in the context of biomization often reflects refined and updated relationship between PFTs and biomes that allows for differentiating between similar types of vegetation e.g. trees vs shrubs vs lianas (Lebamba et al., 2009) or tropical vs non-tropical desert (Lézine et al., 2009). Likewise, the definition of accuracy in the context of the biomization method is qualitative, in that maps of modern vegetation and generated pollen-based PFT maps are superimposed and visually compared (e.g. Lebamba et al., 2009; Prentice et al., 1992; Verlhac et al., 2018).

The biomization method has been the primary tool for investigating modern and past pollen-plant-biome relationships. The method has been successfully used to classify tropical forests in Central Africa (Lebamba et al., 2009) and to capture biome changes in distribution of tropical montane vegetation along a altitudinal transect in West Africa (Verlhac et al., 2018). At a continental scale, the biomization method has successfully classified the majority of vegetation classes from pollen with an exception of the miombo woodlands, tropical dry forests characteristic of central and southern Africa (Jolly et al., 1998). Regional classification of drier biomes using biomization is challenging; while the majority of East Africa vegetation is well classified by the biomization method, misclassifications arise for mosaic open/closed vegetation characteristic savanna biome (Vincens et al., 2006).

From a methodological standpoint, vegetation reconstructions from proxy data via categorical representation may be improved by applying methodologies developed in computer science, such as machine learning. Supervised classification, a branch of machine learning (ML), is a data driven approach where algorithms learn the relationship between data and discrete classes. The field has established a rigorous process for developing, testing, and statistically evaluating and comparing results between different classification models. In particular, the Random Forest algorithm has been shown to provide the most accurate and precise classifications of African biomes at the continental scale from pollen data (Sobol and Finkelstein, 2018). Biomization and other methods can be applied to fossil pollen assemblages to reconstruct past changes in vegetation distribution over time and space to better understand climate-vegetation interactions.

Direct assignment of pollen assemblages to biome categories via machine learning methods has important advantages; the approach preserves the complexity of a biome and produces more objective assignment as it is less reliant on expert knowledge. Furthermore, high processing power and speed of contemporary computing makes it possible to retain entire proxy datasets for analysis and modeling of biomes. Thus, information loss is reduced by using whole multivariate datasets. Resulting models are statistically validated and robust. The trade-off of machine learning pollen-based biome modeling is the potential for the algorithms to identify ecologically irrelevant patterns in the data. Hence, classifications and predictions made by the ML models must be critically evaluated from an ecological perspective.

From a geographical perspective, biome modeling in Africa is spatially biased towards central parts of the continent. As a result, understanding of vegetation changes is restricted to specific biomes, i.e. tropical forests. Although savanna constitutes 60% of sub-Saharan Africa (Scholes and Walker, 1993) it has received relatively limited attention and its classification is challenging. Savanna is the largest contemporary biome in Southern Africa by areal extent (Mucina and Rutherford, 2006, p. 37), yet uncertainty exists about the response of Southern African savanna to ongoing environmental change. Modeling and field-based studies suggest the possibility of a sudden shift of savanna to another stable state driven by rapidly changing environmental conditions (Sala, 2000; Sankaran et al., 2005; Staver et al., 2011; Higgins and Scheiter, 2012). Contemporary ecological evidence indicates that savannas are expanding and invading more open grasslands, a process referred to as bush encroachment (Archer et al., 1995; O'Connor et al., 2014; Skowno et al., 2017; Stevens et al., 2017). Bush encroachment can be attributed to multiple factors including fire suppression, overgrazing, decreasing herbivory, and increasing concentrations of atmospheric CO_2 (O'Connor et al., 2014; Ward, 2005). How past changes and interactions between such factors impacted the direction of savanna shifts, whether towards desertification or bush encroachment, is uncertain. Therefore, classifications of modern savanna vegetation from pollen data must be improved to reliably predict past savanna from fossil pollen sequences.

In summary, vegetation classification from pollen data via categorical representation may be further improved by taking advantage of recent advances in machine learning methodologies, technology, and computing. Incorporating whole sets of multivariate proxy data with rigorously developed machine learning methods is a promising new way to model biomes. The probabilistic approach to biome classification proposed here has the capacity to afford a more nuanced view of paleovegetation, revise current understanding of past biome distribution, and allow for more spatially-detailed reconstructions of vegetation by applying the method to regional and local vegetation classification systems. Furthermore, by increasing spatial resolution to include underrepresented regions and vegetation, such as Southern African savanna, classification performance of models on underrepresented biomes may improve. When applied to fossil pollen sequences from underrepresented regions, such models may provide additional insights into past vegetation dynamics and improve understanding of past biome evolution as well as of future change.

In this paper, we apply a new supervised machine learning method for pollen-based biome classification in Southern Africa. We apply our models to a fossil pollen sequence at site from a transitional zone between savanna and grassland and calculate probabilities of different biomes occurring over the last 60 000 years. We explore potential future responses of the contemporary savanna biome to environmental changes, by reconstructing long-term savanna shifts, linking these shifts to past environmental changes, and placing them in the context of regional paleovegetation reconstructions. Specifically, in this paper we:

1) Present an updated, revised, and expanded modern pollen dataset linked to vegetation zones in Southern Africa that may form the basis for refinement in the future as new data become available;
2) Test the suitability of this dataset for classification of pollen assemblages into biomes and bioregions (using the a priori classification of Mucina and Rutherford (2006)) and prediction of past vegetation states;
3) Develop and validate pollen-based classification models for modern vegetation of Southern Africa;

4) Apply predictive models to a classic fossil pollen sequence for reconstructions of changes in state of past biomes and bioregions, more specifically the Savanna and Grassland biomes;
5) Compare our predictive models' results to previous reconstructions of past vegetation; and
6) Provide open-source R codes for our classification models as templates for ready application to other regions of the world.

2. Materials and methods

2.1. Modern pollen dataset

For training our machine learning classification models, we link modern pollen samples to vegetation classes derived from the classification of Mucina and Rutherford (2006). Building upon previous research, we compile available surface pollen data for Southern Africa, including sites in South Africa, Lesotho, Botswana and Namibia (Fig. 1). Original samples from areas with relatively undisturbed vegetation cover were collected between 1974 and 2014 and represent a variety of substrates (Mucina and Rutherford, 2006; Rutherford and Westfall, 1986).

Pollen counts included in our modern pollen dataset were obtained from several publications (Cooremans, 1989; Jolly et al., 1998; Scott, 1982a, b, 1989; Scott and Cooremans, 1992; Scott et al., 1992), and an on-line database, the African Pollen Database (currently hosted at ftp://ftp.ncdc.noaa.gov/pub/data/paleo/pollen/tiliafiles/apd/). Pollen data from the APD was checked, corrected and revised for inconsistencies by returning to the original count sheets. To further increase the number of data points we added raw pollen counts from new surface samples sites located in semi-arid Southern African savanna (Northern Cape Province, South Africa).

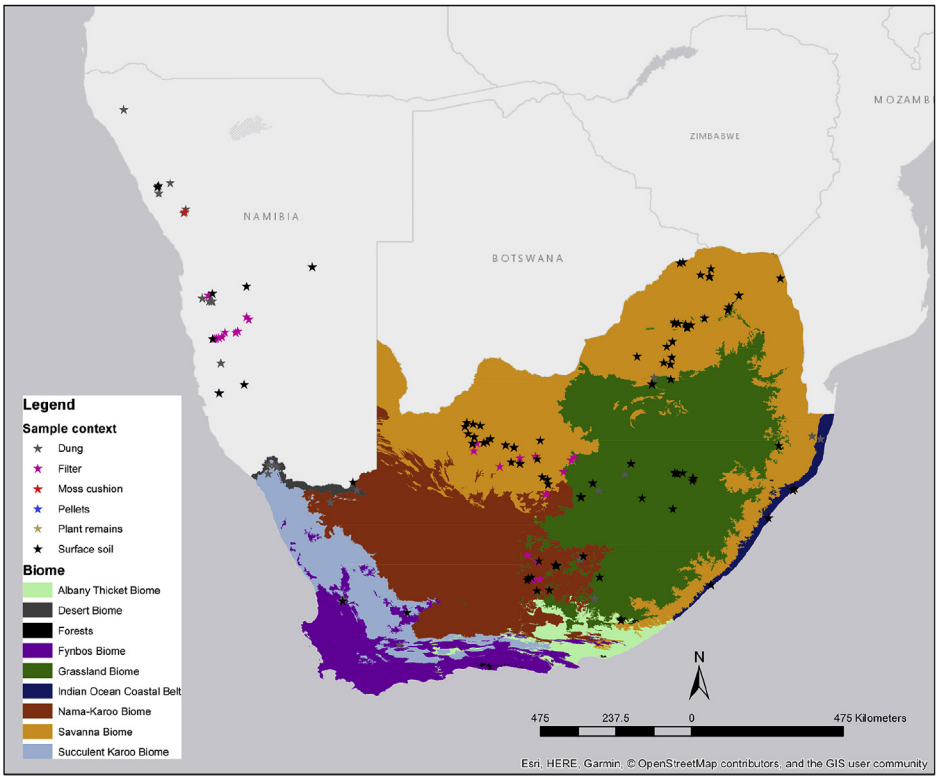
The new modern pollen samples were prepared following a standard procedure for chemical digestion (Fægri and Iversen, 1986) attempting a methodology close to that employed in the published modern pollen samples, which involved use of $ZnCl_2$ solution (S.G. 2) as heavy liquid for separation, and, depending on the matrix material, omission of the exotic *Lycopodium* marker (Stockmarr, 1971), KOH, HF, acetolysis and identifications at magnification of up to 1000x. The new sediment samples were subsampled and 2 tablets of an exotic marker (*Lycopodium*) were added to calculate pollen concentrations (Stockmarr, 1971). Subsamples were treated with 10% hydrochloric (HCl) acid and repeated 10% potassium hydroxide (KOH) baths to remove carbonates and humic acids respectively. Heavy liquid (sodium polytungstate) was used to separate minerals from the pollen fraction (2). To remove fine silicates and other colloidal material not eliminated with the mineral separation, subsamples were then treated with hydrofluoric (HF) acid and followed by acetolysis treatment to clean pollen exine. Safranine was added to stain pollen grains to facilitate identification. Subsamples were mounted in silicone oil on microscope slides and sealed with nail varnish. Pollen grains were identified and counted under Zeiss AX10 Imager.A1 compound microscope, at 400–600× magnification. They were compared with a digital reference collection and published pollen identification resources (Carrión et al., 1999; van Geel, 1978; van Geel and Aptroot, 2006; Scott, 1982b; Sowunmi, 1973, 1995; van Zinderen Bakker, 1953, 1956; van Zinderen Bakker and Coetzee, 1959; van Zinderen Bakker et al., 1970). Pollen was identified to the lowest taxonomic level possible.

2.2. Vegetation classification

For classification of broad-scale vegetation of Southern Africa we assign modern pollen samples to vegetation types including biomes (e.g., Grassland Biome representing grassy veld and Savanna Biome representing woody veld) and subdivisions, their constituent bioregions, following Mucina and Rutherford's (2006) vegetation classification for South Africa, Lesotho and Swaziland (Mucina et al., 2006a, b, c). Biomes of Southern Africa range from the dry Desert Biome on the west coast, characterized by high annual evaporation rates and low soil moisture, to lush coastal forests of the Indian Ocean Coastal Belt on the east coast, where soil moisture is relatively high throughout the year. Most of Southern Africa, however, is dominated by the three largest biomes that receive summer rainfall – the Savanna, Grassland, and Nama Karoo Biomes (Mucina and Rutherford, 2006). Characterized by the coexistence of a grass layer and woody vegetation, the Savanna Biome varies in the relative proportion of trees ranging from open parklands to denser woodlands (Rutherford, 1982; Scholes and Archer, 1997). On the other hand, the cool and moist vegetation of the temperate frost-controlled Grassland Biome dominates high altitude summer-rain areas, such as the high central plateau (the Highveld), the high Lesotho and Drakensberg Mountains. The landlocked Nama Karoo extends over the southern central plateau of South Africa and southeastern Namibia. The dominant vegetation is a mix of xeric dwarf shrubs and grasses with arboreal species confined to riparian zones (Palmer and Hoffman, 1997). Precipitation, seasonality, and soil conditions that vary seasonally or regionally are the main factors that determine the relative proportion of shrubs and grasses showing an inverse relationship between shrubs and rainfall (Palmer and Hoffman, 1997).

The Southern African vegetation classification is based on climate variables, physical terrain, as well as processes operating at regional scale, such as microclimate, substrate, topography or disturbance (Mucina et al., 2006b, pp. 40–51). The system defines biomes as the largest terrestrial vegetation communities identified at a continental or sub-continental scale characterized by ecologically important climate parameters, but excludes anthropogenic zones (Mucina et al., 2006, pp. 31–40). Using climate variables, e.g. mean annual precipitation (MAP), mean annual temperature (MAT), and mean frost days (MFD), Southern African vegetation is classified into discrete groups using the Classification and Regression Trees (CART) method (Mucina et al., 2006b, pp. 38–40). For instance, at a subcontinental scale Savanna biome is characterized by MAP of 495 mm, MAT averaging 18.7 °C, and estimated 16 MFD per year (Mucina et al., 2006b, p. 40). Relative to the Savanna Biome, Grassland is characterized by higher MAP (661 mm), lower MAT (14.7 °C), and severe frost incidence (40 MFD) (Mucina et al., 2006b, p. 40). Bordering with both Savanna and Grassland is a dry shrubland of the Nama Karoo Biome characterized by lowest MAP (208 mm), intermediate MAT (16.3 °C), and frost incidence (35 MFD) similar to Grassland (Mucina et al., 2006b, p. 40). The Indian Ocean Coastal Belt (IOCB) Biome on the eastern coast of South Africa is characterized by high precipitation (MAP 985 mm) and warm temperatures (MAT 20 °C; 0 MFD) that support subtropical coastal forests (Mucina et al., 2006b, p. 40).

Regional variation in precipitation and temperature leads to distinctions in classification of bioregions which consist of more specific vegetation categories (Mucina et al., 2006b, pp. 48–49). For instance, the Central Bushveld and the Eastern Kalahari Bushveld are two climatically distinct types within the broader Savanna Biome (Rutherford et al., 2006). The Central Bushveld is characterized by higher precipitation (MAP 559 mm) with higher temperatures (MAT 18.4 °C), and lower frost incidence (13 MFD) representing a sub-humid savanna sub-type. On the other hand, while Eastern Kalahari Bushveld has lower precipitation (MAP 362 mm), and lower temperature (MAT 17.8 °C), and higher incidence of frost (33 MFD) representing an arid savanna subtype (Rutherford et al., 2006). Similarly, bioregions within the Grassland



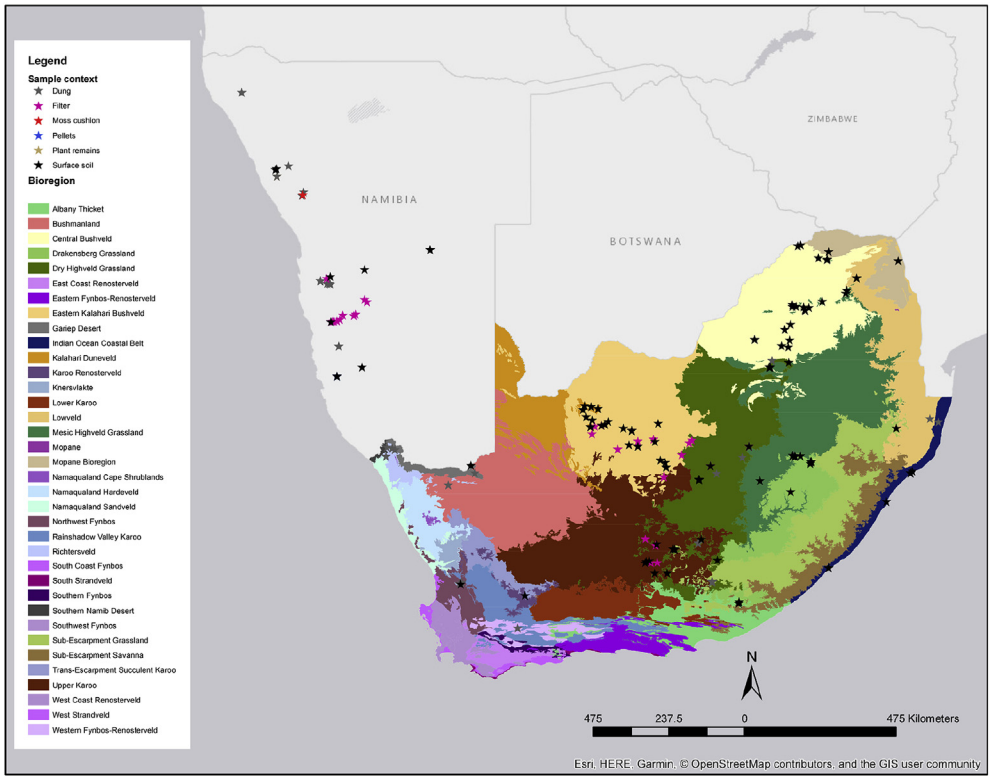


Fig. 1. Vegetation maps of Southern African vegetation units biomes (top) and bioregions (bottom) showing the location of modern pollen samples used for building the modern classification models.

biome may be also be distinguished on the basis of altitude- or frost-driven climate and vegetation composition. For instance, the endemic-rich vegetation dominated by C_3 grasses of the Drakensberg Grassland Bioregion occupies high altitude areas of the Drakensberg mountain range where precipitation is high (MAP 732 mm), and low temperature (MAT 10.8 °C) results in high frost incidence (78 MFD). On the other hand, the Sub-Escarpment Grassland Bioregion, found at a lower altitude of the Drakensberg foothills, has similar precipitation (MAP 763 mm) but warmer temperatures (MAT 15.5 °C) and lower frost incidence (21 MFD) (Mucina et al., 2006a). As such, Southern African vegetation representative of a particular biome or bioregion is linked to specific climatic conditions.

2.3. Fossil pollen sequence

For prediction and reconstructions of past vegetation states, we chose a fossil pollen record that occurs in one of the regions that is best represented by the modern pollen record. Located in a transitional zone between grassland and savanna systems, Wonderkrater (24° 25′ 47.5" S, 28° 44' 37.5" E) is one of the longest pollen sequences, covering the last 60 000 years and one of the most complete terrestrial records from Southern Africa (Scott 1982a, 2016). The site is located in the Limpopo Province of South Africa and represents contemporary sub-humid sub-type in the Savanna Biome in an area where different woodland types converge (Scott, 1982a, 2016; Rutherford et al., 2006). It is situated in a sub-humid savanna type (Central Bushveld) on the northwestern edge of the Sprinbokvlakte Thornveld bordering on the Central Sandy Bushveld and lies close to the Waterberg Mountain Bushveld (Rutherford et al., 2006; Scott, 2016).

While the pollen record from this thermal spring site has been the subject of previous palaeoenvironmental reconstructions by various methods, environmental conditions during and following the Pleistocene-Holocene transition at Wonderkrater remain unclear. Specifically, interpretations of paleo-precipitation during the early Holocene are in disagreement, with some studies showing increased precipitation (Chevalier & Chase, 2015, 2016; Truc et al., 2013), and others suggesting relatively dry conditions (Scott, 1982a, 2016; Scott et al., 2003; Scott and Thackeray, 1987). The ML approach does not attempt to provide quantitative reconstruction of climate variables. Rather, it shows probabilities of biomes. Thus, the method will provide complementary information to help evaluate the results of the quantitative studies.

For prediction of past biomes from fossil pollen data, we combine four adjacent cores (Boreholes 1 to 4, or B1-4) from Wonderkrater. Apart from one marked hiatus and some minor hiatuses, the cores collectively represent the last 60 ka BP (Scott, 1982a, 2016). For the chronology of the combined sequence, we rely on the latest published scheme by Scott (2016). For the entire B3 and the upper part of B4, the age model is based on radiocarbon dating and was developed using Baysian age depth modeling program, Bacon (Blaauw and Christeny, 2011). B3 was collected by L. Scott and colleagues in 1974, and represents the most recent period from present to 19.6 cal ka BP (Scott, 1982a, 2016). B1 and B2 were collected by van Zinderen Bakker in 1971 and, combined, represent the middle part of the sequence, an interval estimated between ~25 and 27 cal ka BP (Scott, 2016). L. Scott and colleagues collected B4 in 1977. The top of the core is dated to 7.7 cal ka BP and the lower part extends back to an estimated 60 ka BP; the lower age bound is inferred from extrapolations from radiocarbon age models, plus optically stimulated luminesce (OSL) and infrared stimulated luminescence (IRSL) dating (Backwell et al., 2014; Scott, 2016). B4 has a hiatus between 30 and 41 cal ka BP (Scott, 2016) and an overlaps with B3 for c. 10 000 yr (between 7 and 8 cal ka BP).

We consider the chronology of the single core B3 (up to *c.* 18.5 cal ka BP) and overlapping section to be relatively reliable, while the chronology for B1 and B2 (*c.* 25–27 cal ka BP), and B4 (*c.* 41–60 cal ka BP) is progressively less reliable in view of the assumptions introduced through extrapolation and correlation. An alternative, younger age model for the older section is presented in Chevalier and Chase (2015, 2016). However, we opted to use the age model from Scott (2016) as it recognises the high probability of a large erosional hiatus in the sequence associated with a prominent sand layer (Backwell et al., 2014; Scott, 1982a). The application of ML biome classifications to this fossil sequence does not directly depend on the age model. We acknowledge there may be some uncertainty in the age models; however, this does not impinge upon our interpretations of the reconstructed biomes. These interpretations are mainly focussed on the Late Pleistocene and Early Holocene, time periods for which there are considerably more confidence in the age models.

For training of classification models, we harmonize the Wonderkrater fossil pollen data to the modern pollen dataset. Any fossil pollen taxon that did not correspond to modern dataset was reassigned to the next highest taxonomic classification. For instance, fossil *Ascolepsis* was reassigned to its parent family Cyperaceae. Where re-assignments of fossil taxa to the highest taxonomic levels were impossible, given fossil taxa were removed from the dataset and their abundances excluded from the total pollen sum. Aquatic taxa do not appear in the modern pollen assemblages and are not expected to be characteristic of Southern African biomes and bioregions. As a result, aquatic taxa from Wonderkrater pollen sequence (*Eriocaulon*, *Hydrocotyle*, Nymphaea, and Potamogetonaceae) were removed. Similarily, Lentibulariaceae are often aquatic taxa and were therefore also excluded. Furthermore, Palmae, *Myrothamnus, Brachystegia*, *Canthium*, and *Encephalartos* are fossil pollen taxa present in the Wonderkrater sequence that were removed on the ground of low counts (<10 grains) and/or limited ecological distribution. For instance Palmae pollen is an indicator of warm frost-free conditions and not found in grasslands (Scott, 1982b). Likewise, *Myrothamnus* is a prominent element of arid environments of the Namib Desert margin (Scott, 1982b).

To investigate latent structure in the pollen data as a function of vegetation from South Africa, Lesotho and Swaziland we perform ordination by Detrended Correspondence Analysis (DCA) (Hill and Gauch, 1980). The ordination was conducted in R version 3.4.3, using the *decorana* function in the *vegan* package version 2.4–6 (Oksanen et al., 2017).

2.4. Machine learning models for classification of modern southern African vegetation

For our models' development, we restrict the modern pollen data to Southern Africa, Lesotho, and Swaziland for correspondence with the vegetation classification system (Mucina and Rutherford, 2006), which lacks classification for Namibia (the whole data is available at GitHub). We assign biome and bioregion labels to the modern pollen data points by projecting them onto the vegetation maps in ArcGIS and extracting class labels for each data point. Thus, for the development of the classification models we exclude 47 modern pollen surface samples from the original dataset that are located in Namibia and fall outside the vegetation classification system (Mucina and Rutherford, 2006). The remaining 164 surface pollen samples were used for training biome and bioregion models.

To predict past biomes from fossil pollen data we first train and validate classification models using modern pollen data and modern vegetation classification system. To our regional dataset we adapt a previously developed method using the Random Forest algorithm (Sobol and Finkelstein, 2018). We build the classification

model using *randomForest* package for R version 4.6–12 (Liaw and Wiener, 2002) and optimize it using the *tuneRF* function in the same package. We use bootstrap aggregating to subsample pollen data for training. The out-of-bag (OOB) data are used to estimate the class error and measure the importance of each variable (Simpson and Birks, 2012). We optimize the biome and bioregion models by i) varying the number of sites representing vegetation units (0, ≤4, ≤8, ≤10); ii) varying the number of decision trees in the forest; and iii) finding an optimal number of the random subset of predictor variables/pollen taxa by setting $mtry = best$ in the *tuneRF* function.

We assess our models' performances on the OOB set using accuracy as the primary evaluation metric. Model accuracy is the model's ability to correctly assign a given vegetation unit as compared to the known vegetation classes; i.e. number of correct model classifications for a given vegetation unit divided by the total number (reported as percent). The model with the lowest OOB error rate was chosen and applied to the fossil pollen sequence to reconstruct past vegetation change. Additionally, to examine the models' classification of individual vegetation classes we calculate the following evaluation metrics: recall, precision, F1 and kappa statistics. Lastly, we calculate the importance of each pollen taxon for both models through the Mean Decrease in Accuracy (MDA); the difference in accuracy of a model trained on all pollen data versus the accuracy of a model trained with a given taxon randomized. By systematically randomizing each pollen taxon, the signal associated with each of them is removed to reveal the contribution of each taxon to model performance (Cutler et al., 2007).

2.5. Application of the classification models to the Wonderkrater fossil pollen records for prediction of past vegetation

Once trained and validated on the modern pollen and vegetation data, we apply the biome and bioregion models to the Wonderkrater fossil pollen sequence for predictions of past vegetation. To predict past changes in biomes and bioregions at Wonderkrater, we calculate the probability of different vegetation classes occurring through time at Wonderkrater. As the Wonderkrater sequence combines four individual pollen records (B1–B4), we predict biomes and bioregions on the combined single sequence. To measure non-parametric rank correlation between vegetation classes predicted for Wonderkrater, we calculate Kendall's rank correlation coefficient (τ) as the distribution of the vegetation classes is not normal.

3. Results

3.1. Modern pollen dataset

The updated modern pollen dataset consists of 211 samples primarily obtained from surface soil, with a small number from dung, middens, and pollen-trap filters. We obtain 68 modern pollen assemblages from literature and another 121 from the African Pollen Database. These database records were checked and compared against original counting sheets; all discrepancies were corrected to the values from the original counting sheets. We also present 22 previously unpublished raw pollen counts from Namibia and South Africa including Eastern Cape, Western Cape, Northern Cape, and Limpopo Provinces.

The dataset captures broad climate gradients ranging from arid to subtropical representing a regional vegetation gradient and from desert, xeric shrubland, savanna, grassland, subtropical forests to Mediterranean shrubland ('fynbos') biomes. Specifically, surface pollen samples represent the following Southern African biomes (Mucina and Rutherford, 2006): Savanna (67), Grassland (53), Nama Karoo (18), The Indian Ocean Coastal Belt Biome (11), Desert (5), Fynbos (5), Succulent Karoo (4) and Albany Thicket (1). Another 47 surface samples from the western coast of Namibia may be classified as belonging mostly to other unspecified arid or semi-arid biomes similar to Desert, Nama Karoo, and Savanna (Irish, 1994). A total of 157 pollen taxa are represented in the updated modern pollen dataset. On average, the pollen count per sample is 322 grains with a minimum of 100 grains and a maximum of 2699 grains.

The distribution of biomes (Fig. 2a) and bioregions (Fig. 2b) as function of associated pollen assemblages in the ordination space is characterized by an overlap between different vegetation classes (Code A.1). Nonetheless, some distinct groupings of vegetation classes may be observed. For example, there is no overlap between the Desert and Indian Ocean Coastal Belt Biomes or between Nama-Karoo and Indian Ocean Coastal Belt. In fact, the Desert and Indian Ocean Coastal Belt Biomes plot on the opposite side of DCA Axis 1 (Fig. 2a). Similar groupings are also observed in regional classification of vegetation (Fig. 2b). For instance, Savanna may also be separated into distinct sub-types despite some overlap.

3.2. Pollen classification models of modern southern African vegetation

The following parameters yielded biome and bioregion models with lowest out-of-bag (OOB) error rate. Both models were run 100 times and in both cases we grew 500 trees in the forest with the optimal number of the random subset of predictor pollen taxa at each split being 11. The minimum number of pollen sites representing each biome was chosen to be 8, while the minimum number of pollen sites representing each bioregion was chosen to be 10; biomes and bioregions represented by fewer sites were not considered in the model development.

The selected surface pollen samples represent four biomes (Fig. 1, Table S.1) and seven bioregions (Fig. 1, Table S1). Pollen samples within the Savanna Biome (60) fell into two bioregions: Central Bushveld (21) and Eastern Kalahari Bushveld (39). Similarly, modern pollen samples from Grassland (44) fell into three bioregions: Drakensberg Grassland (10), and Dry Highveld Grassland (23), and Sub-Escarpment Grassland (11). Modern pollen samples from Nama Karoo (18) are represented by only one bioregion with >10 sites, the Upper Karoo (16) (Table S1). The Indian Ocean Coastal Belt Biome does not have bioregional subdivisions (Mucina et al., 2006d), thus we retain the Indian Ocean Coastal Belt (11) label for the bioregional classification. A pollen diagram shows the modern pollen-biome and bioregion relationships of modern pollen data used in our analyses (Fig. 3). The results of classification performance for our modern biome (Table 1, Table S2) and bioregion (Table 2, Table S3) models are presented in the standard form of confusion matrices.

The overall performance of the modern biome model is high with the out-of-bag (OOB) estimate of error rate of 16% with Kappa statistic equal to 0.85 i.e. the model correctly predicts biome based on pollen assemblages 85% of time (Table 1). For individual biomes, we achieve highest performance based on recall for Savanna (96%), Grassland (89%), and Indian Ocean Coastal Belt (82%).

The overall performance of the bioregion model is relatively high with the OOB estimate of error rate of 22% and Kappa statistic equal to 0.73 (Table 2). For individual bioregions, we achieve the highest classification performance based on the recall metric for Central Bushveld (95%) and Eastern Kalahari Bushveld (95%), Drakensberg Grassland (90%), and Indian Ocean Coastal Belt (91%) bioregions. In addition, a high classification is achieved for Sub-Escarpment Grassland (73%).

The variable plots show the contribution of pollen taxa to

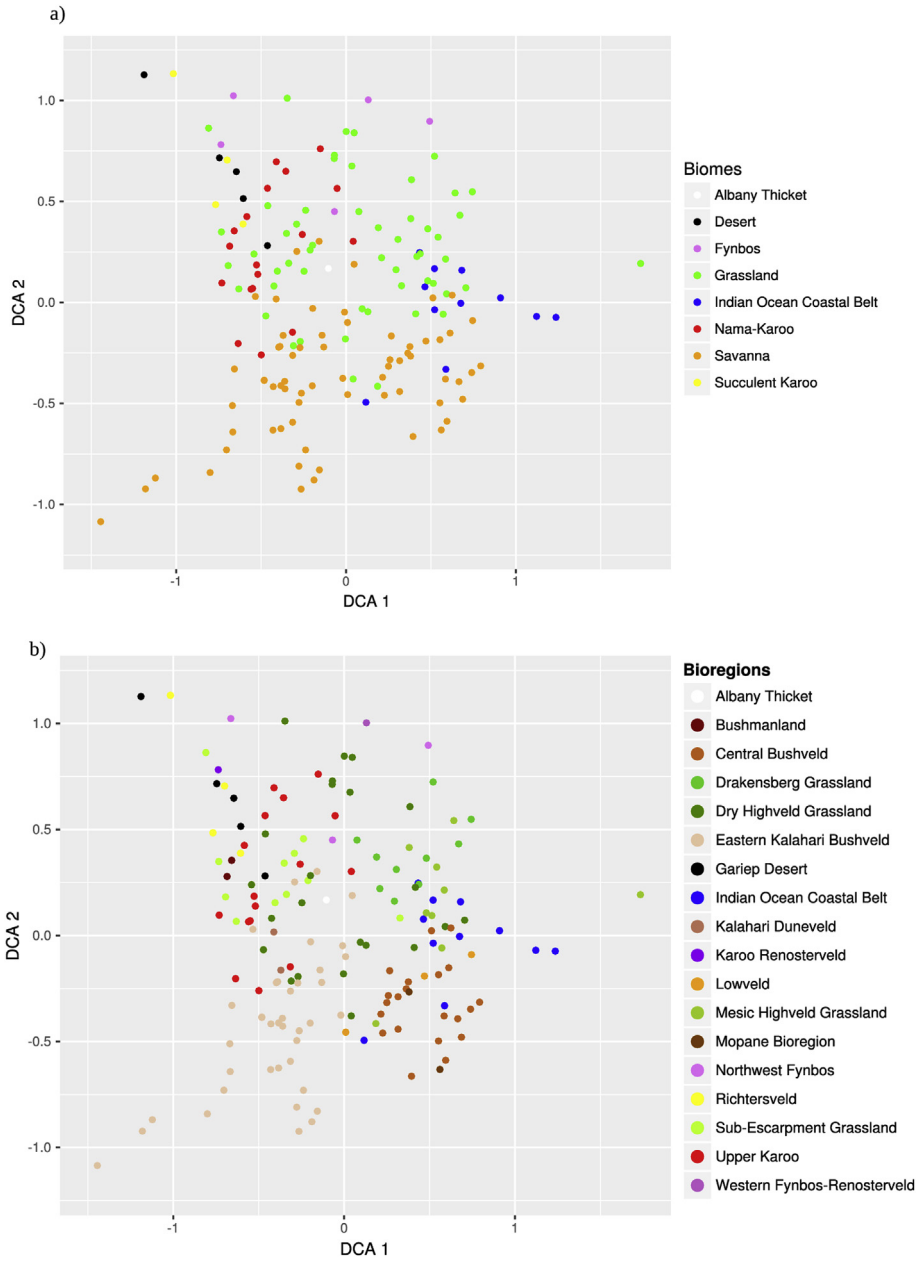


Fig. 2. Ordination by Detrended Correspondence Analysis (DCA) of modern pollen sites (axes 1 and 2 shown). Pollen samples are color-coded according to their corresponding vegetation units a) biomes and b) bioregions (Rutherford and Mucina, 2006). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

accurate classification in the biome (Fig. 4a) and bioregion (Fig. 4b) models (Table S.4). For classification of modern biomes, undifferentiated Amaranthaceae including Chenopodiaceae (Scott, 1982b) is the most important taxon contributing 2.6% towards the accurate classification of modern biomes (Fig. 4a). Other groups of important pollen taxa include *Tarchonanthus* type, *Artemisia afra* type (Scott, 1982b, 1989, 2016), and undifferentiated Asteraceae, each contribute 2% to the biome model. Undifferentiated Combretaceae (3.8%), is the most important taxon for predicting and classifying Southern African bioregions. Undifferentiated Amaranthaceae (3.1%) and *Tarchonanthus* type (3.1%) are also important taxa. To a lesser extent, a group of pollen taxa comprising *Burkea africana* (2.5%), undifferentiated Asteraceae (2.3%), *Pinus* (2%), and *Anthospermum* (1.8%) also contribute to accurate classification of Southern African bioregions.

3.3. Prediction of past biomes and bioregions from a fossil pollen record. presentation of data

Traditional classification approaches typically categorize data into discrete groups. Our machine learning models are no different in the sense that they classify modern pollen assemblages into discrete vegetation units, i.e. biomes and bioregions. However, when applied to fossil pollen sequences for reconstructions of past vegetation states, the method provides a more probabilistic view of the likelihood of vegetation units occurring in the past. For instance, suppose that at a given time savanna probability reaches 70% while grassland probability is 25% with other biomes having a

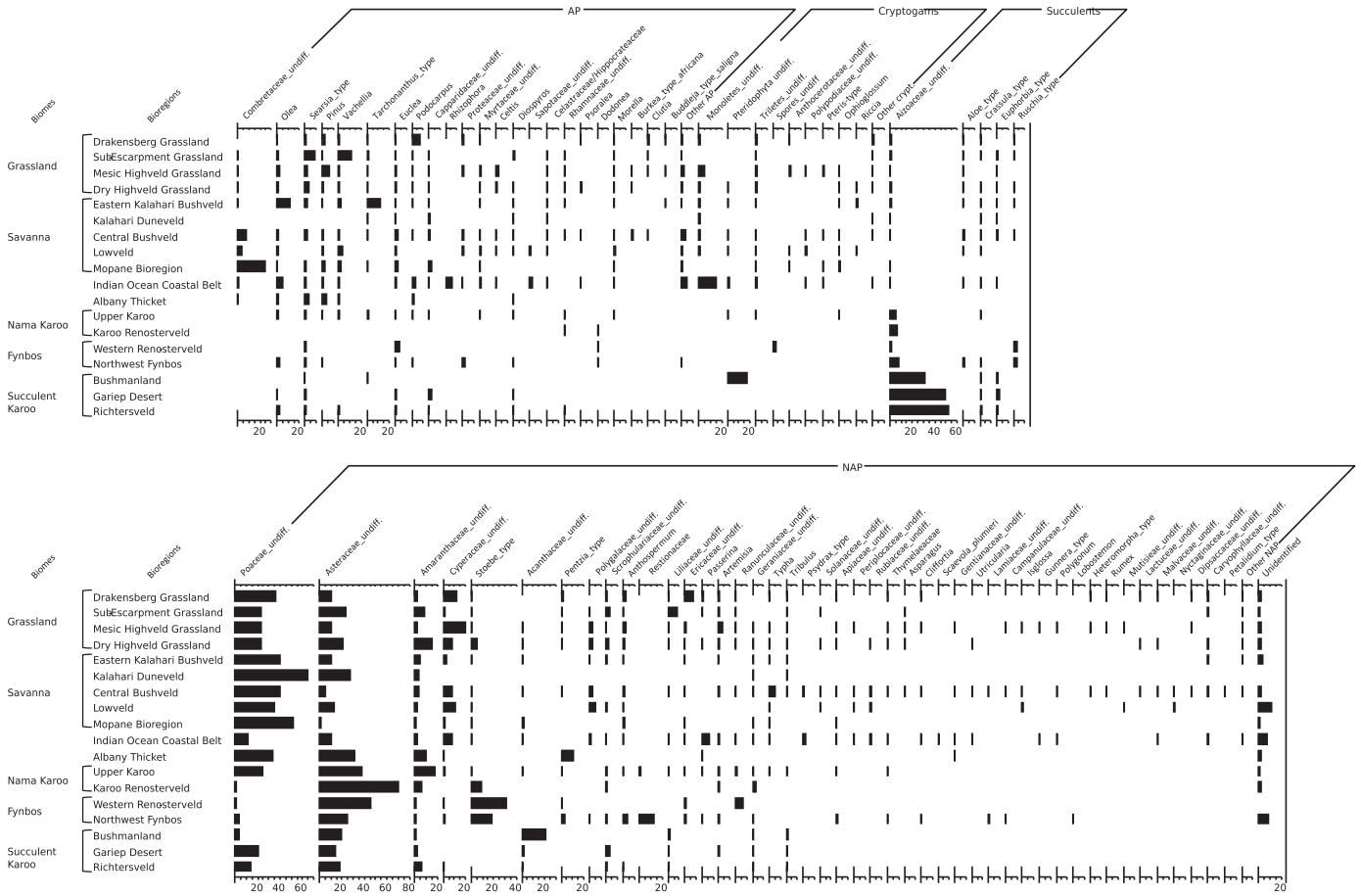


Fig. 3. Summary pollen diagram of 164 modern pollen spectra from South Africa. Taxa that totaled less than 1% were combined as “Other AP”, “Other crypt” or “Other NAP”.

Table 1
Confusion matrices showing performance of Random Forest model for classifying southern African modern biomes using modern pollen assemblages. Evaluation metrics for classification of individual biomes were calculated on the OOB error. Number of correct predictions run diagonally and are highlighted in bold. Recall for each biome type is calculated as the number of correct classifications for a given known biome. Precision for each classified biome is calculated as the proportion of correctly classified biome to the sum of all classifications.

Model OOB error 0.84 Model Kappa 0.85		Predicted biomes			Evaluation metrics				
		Grassland	IOCB	Savana	OOB error	Recall	Precision	F1	Individual Kappa
True Biomes	Grassland	47	0	6	0.11	0.89	0.94	0.91	0.82
	IOCB	0	9	2	0.18	0.82	1.00	0.90	0.74
	Savana	3	0	64	0.04	0.96	0.89	0.92	0.80

Table 2
Confusion matrices showing performance of Random Forest model for classifying southern African modern bioregions using modern pollen assemblages. Evaluation metrics for classification of individual bioregions were calculated on the OOB error. Number of correct classifications run diagonally and are highlighted in bold. Recall for each bioregion class is calculated as the number of correct classifications for a given known bioregion. Precision for each classified bioregion is calculated as the proportion of correctly classified bioregion to the sum of all classifications.

Model OOB error 0.22 Model Kappa 0.73		Predicted bioregions						Evaluation metrics					
		CB	DG	DHG	EKB	IOCB	SEG	UK	OOB error	Recall	Precision	F1	Individual Kappa
True bioregions	Central Bushveld	**20**	0	0	0	1	0	0	0.05	0.95	0.83	089	0.87
	Drakensberg Grassland	0	**9**	0	1	0	0	0	0.10	0.90	1.00	0.95	0.94
	Dry Highveld Grassland	2	0	**9**	7	0	0	5	0.61	0.39	0.56	0.46	0.39
	Eastern Kalahari Bushveld	0	0	1	**37**	0	0	1	0.05	0.95	0.77	0.85	0.76
	Indian Ocean Coastal Belt	1	0	0	0	**10**	0	0	0.09	0.91	0.91	0.91	0.95
	Sub-Escarpment Grassland	1	0	2	0	0	**8**	0	0.27	0.73	0.89	0.80	0.84
	Upper Karoo	0	0	4	3	0	1	**8**	0.50	0.50	0.57	0.53	0.47

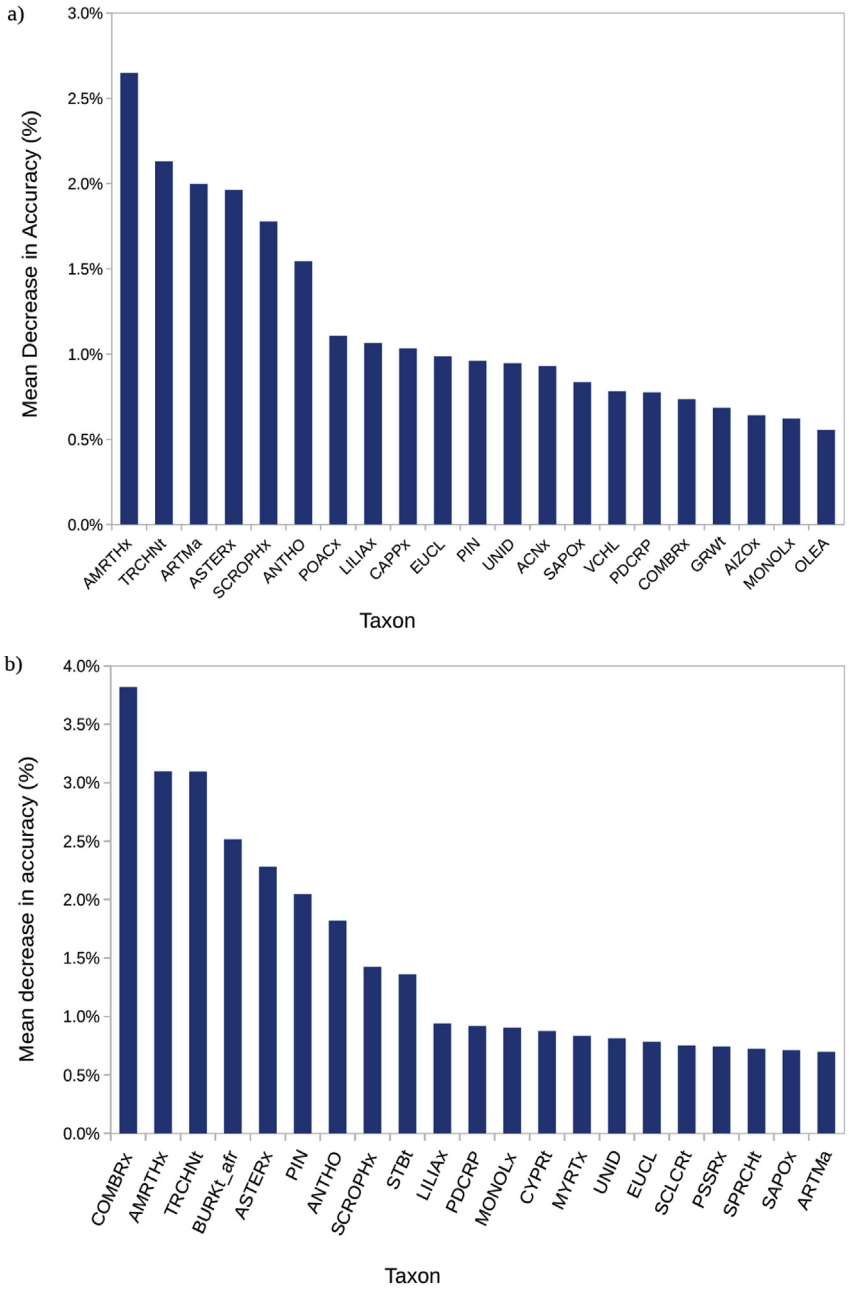


Fig. 4. Mean Decrease in Accuracy (MDA) calculated for a) biome and b) bioregion models showing pollen taxa that contribute to high classifications. For each model the most important 21 taxa are plotted. Abbreviations of pollen taxon names along with their MDA percentages may be found in (Table S.4).

collective probability of 5%. In this case, the relative probabilities of different biomes indicate an environment belonging to a savanna biome. A grassland biome is less likely and representation from other biomes, of which the curves only mean they share a few pollen taxa, is insignificant. Equal probabilities of savanna and grassland would be consistent with a more open canopy savanna environment, while higher probabilities of the latter would be interpreted as an open environment with a reduced tree cover, potentially representing a grassland biome. Thus, in contrast to the binary nature and biases of traditional classification approaches, our method offers a more complex and nuanced view of past vegetation dynamics. The results are measures of vegetation type and not climatic factors like temperature and moisture, which can only arbitrarily be estimated from their ranges within a biome according to its definition.

3.4. Wonderkrater

As a result of the harmonizing process of the raw Wonderkrater fossil pollen data (Scott, 2016, supplementary material) to our modern pollen dataset, 20 pollen taxa were reassigned and 8 taxa were excluded (Table S.5). By applying our modern models to the Wonderkrater fossil pollen sequence we are able to estimate probabilities of different biomes (Fig. 5) and bioregions (Fig. 6) occurring over time at Wonderkrater. There is a negative correlation between Savanna and Grassland as well as Savanna and Nama Karoo biomes while no significant correlations exists between other biomes or bioregions predicted at Wonderkrater (Table 3).

At sub-continental scale, the Savanna and Grassland Biomes are the top two biomes that most likely occurred at various times over the last 60 k at Wonderkrater (Figs. 4 and 5; Table S2 and S3) with

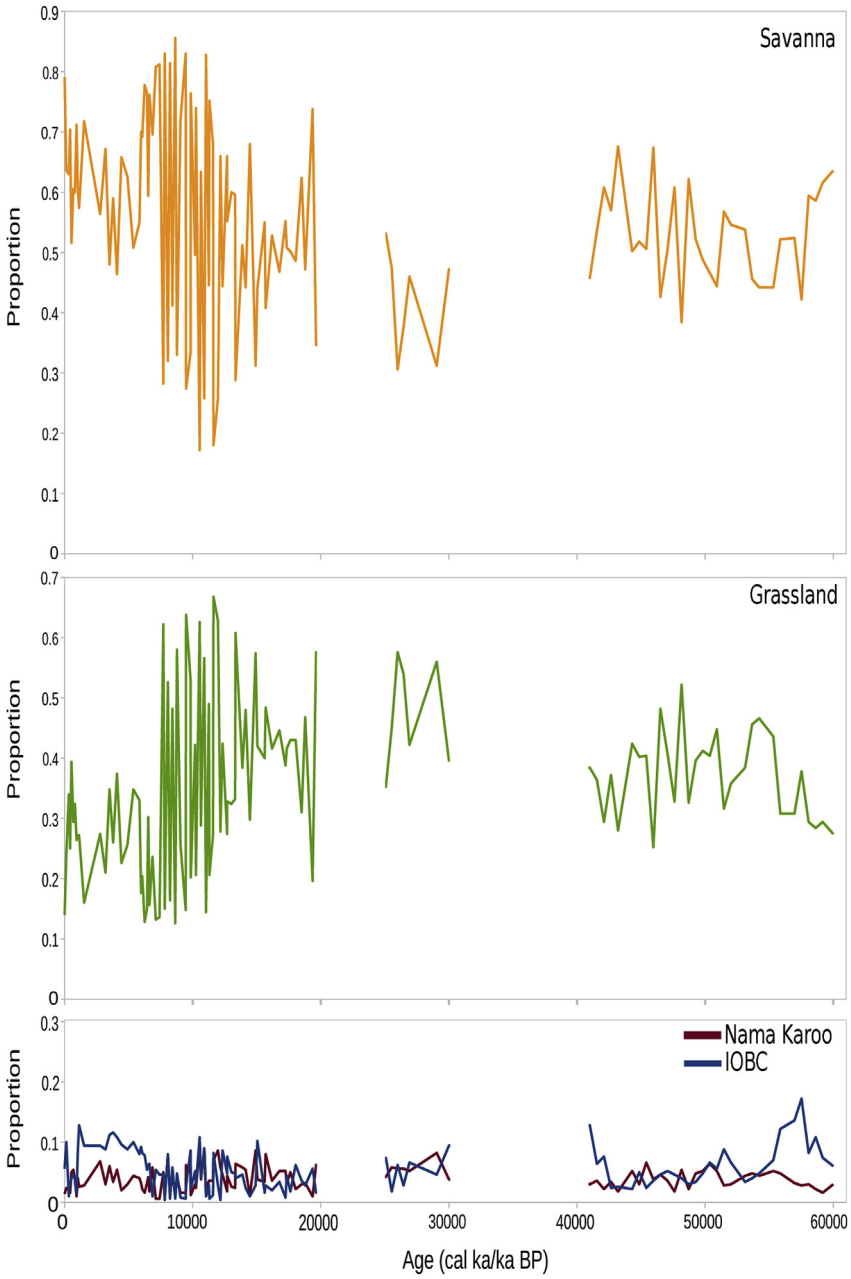


Fig. 5. Models predictions of changes in past probabilities of Southern African biomes (Mucina and Rutherford, 2006) for Wonderkrater. The acronym IOCB refers to Indian Ocean Coastal Belt.

the Indian Ocean Coastal Belt and the Nama Karoo Biomes having relatively low predicted probabilities of occurring at Wonderkrater. A prominent feature of the predicted biomes is the negative seesaw relationship between Savanna and Grassland Biomes (Fig. 5). From present to 7 cal ka BP the most probable, but variable, biome predicted at Wonderkrater is Savanna. Variable probability is also predicted for Grassland during this period, while the probability of Indian Ocean Coastal Belt shows a sustained increase between 1 and 6.3 cal ka BP but never enough to assume that this biome was likely to have occurred. The most frequent and amplified oscillation between the predicted probabilities of Savanna and Grassland Biomes are observed between 7.7 and 15.7 cal ka BP. During this period Grassland becomes the dominant biome prediction at several times, most notably between 11.6-11.9 and 9.5–9.8 cal ka BP. Although Savanna is predicted to be dominant biome for the period between 16 and 20 cal ka BP, the proportions of Savanna and Grassland are relatively equal.

Predictions for the Wonderkrater record shows increasing proportions of Grassland between 25 and 30 cal ka BP suggesting Grassland as most likely biome assignment followed by a hiatus. The record resumes at *c.* 41 ka BP with Savanna as mostly the dominant type of vegetation until the sequence terminates. However, a see-saw trend in the predicted probabilities of Savanna and Grassland Biomes shows that the two biomes reach relatively equal values particularly during the period between 46.5 and 54 ka BP when the probability of grassland exceeds that of savanna.

Likewise, regional vegetation at Wonderkrater has not changed significantly over the last 60 000 year as indicated by our bioregional model (Fig. 5). The most probable bioregion predicted for Wonderkrater, is the Central Bushveld with probabilities ranging

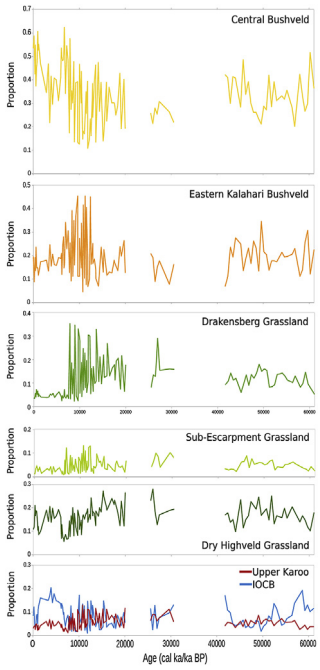


Fig. 6. Models predictions of changes in past probabilities of Southern African bioregions (Mucina and Rutherford, 2006) for Wonderkrater. The acronym IOCB refers to Indian Ocean Coastal Belt.

Table 3
Pairwise comparison of correlation coefficients (Kendal's tau) measuring the statistical relationship between reconstructed biomes (left) and bioregions (right) at Wonderkrater.

Pairwise Biomes	τ	P-value	Pairwise Bioregions	τ	P-value
SV + G	−0.84	<0.01	CB + DG	−0.56	<0.01
SV + NK	−0.60	<0.01	CB + DHG	−0.45	<0.01
SV + IOCB	−0.15	0.02	CB + EKB	0.04	0.51
G + NK	0.56	0.00	CB + IOCB	0.08	0.20
G + IOCB	−0.01	0.87	CB + SEG	−0.45	<0.01
NK + IOCB	0.02	0.79	CB + UK	−0.58	<0.01
			DG + DHG	0.36	<0.01
Biome acronym denote: SV Savana,			DG + EKB	−0.31	<0.01
G Grassland, NK Nama Karoo,			DG + IOCB	−0.13	0.04
IOCB Indian Ocean Coastal Belt			DG + SEG	0.54	<0.01
			DG + UK	0.61	<0.01
			DHG + EKB	−0.11	0.09
			DHG + IOCB	−0.07	0.27
			DHG + SEG	0.31	<0.01
			DHG + UK	0.51	<0.01
			EKB + IOCB	−0.21	<0.01
			EKB + SEG	−0.32	<0.01
			EKB + UK	−0.35	<0.01
			IOCB + SEG	−0.13	0.05
			IOCB + UK	−0.04	0.50
			SEG + UK	0.52	<0.01
			Bioregion acronyms denote: CB Central Bushveld, DG Drakensberg Grassland, DHG Dry Highveld Grassland, EKB Eastern Kalahari Bushveld, IOCB Indian Ocean Coastal Belt, SEG Sub-Escarpment Grasslan, UK Upper		

between 63% at 6.6 cal ka BP and 11% at 11.6 cal ka BP; followed by the Eastern Kalahari Bushveld Bioregion ranging between 45% and 5%, and two grassland bioregions: Drakensberg Grassland and Dry Highveld Grassland, with probabilities ranging 2%–35% and 6% to 28% respectively. The remaining bioregions have variable probabilities but none exceed 30% (Fig. 5, Table S.4).

Overall, the trend in the probabilities of bioregions is more variable as compared to biomes, particularly in the fluctuation of the dominant bioregion, the Central Bushveld. Between 0 and 7 cal ka BP, Central Bushveld Bioregion (Fig. 5), although variable, generally dominates the regional vegetation at Wonderkrater. Between 7 and 20 cal ka BP our bioregional model predicts a highly variable see-saw trend in the probability of all bioregions. Although the Central Bushveld Bioregion remains the most probable during much of this period, Eastern Kalahari Bushveld, along with Drakensberg and Dry Highveld Grasslands are predicted to dominate several times particularly between 9.5 and 12.5 cal ka BP (Fig. 5). Between 13 and 19 cal ka BP the frequency and magnitude of the see-saw trend decrease. From 40 to 60 cal ka BP the Central Bushveld Bioregion is predicted to fluctuate somewhat but remain the most probable bioregion with varying probabilities of the Eastern Kalahari Bushveld, Dry Highveld Grassland, and Drakensberg Grassland bioregions.

4. Discussion

4.1. Classification of modern pollen assemblages into vegetation units

In the ordination space, modern pollen assemblages are characterized by overlap between sites from various biomes (Fig. 2a) and bioregions (Fig. 2b). This overlap suggests that the dataset captures a continuum of plant communities spanning broad transitional zones. Arid biomes such as Desert and Succulent Karoo plot on one end of the DCA Axis 1 while the moist Indian Ocean Coastal Belt Biome plots on the opposite end (Fig. 2a). Similarly, Savanna bioregions are distributed along Axis 1 with the moist Savanna type (Central Bushveld) plotting closer to Indian Ocean Coastal Belt and the dry Savanna type (Eastern Kalahari Bushveld) plotting closer to the arid vegetation. Thus, we interpret the DCA Axis 1 to approximate a moisture gradient. The DCA Axis 2 separates Savanna and Indian Ocean Coastal Belt from the rest of the biomes. Dominant vegetation type in both biomes is a woody cover ranging from open savanna to denser forests. Hence, Axis 2 may represent the divide between denser and more open vegetation most likely driven by a combination of factors such as herbivory, fire, and frost.

The results of the DCA ordination may serve as a guide concerning our classification models. From the ordination, we can discern vegetation types that may be characterized and predicted on the basis of their pollen composition. For instance, despite some overlap, the Savanna and Grassland biomes are generally distinct in the ordination space (Fig. 2a) and our model achieves high classification scores for these biomes (Table 1). On the other hand, all modern pollen samples from Nama Karoo overlap with samples from Grassland and Savanna in the ordination space. In this case, our model correctly classifies pollen assemblages from the Nama Karoo only 50% of time confusing it more with Grassland and to a lesser degree with Savanna (Table 1).

Likewise, for regional vegetation the ordination may be used to identify bioregions that help to classify and predict bioregions. With the Savanna biome, the Eastern Kalahari Bushveld and Central Bushveld bioregions cluster apart in ordination space and are also classified by our model to the highest accuracy out of all bioregions (Table 2). Similarly, the Sub-Escarpment and Drakensberg Grassland Bioregions plot separately in the ordination space and our model classifies them to a relatively high accuracy. On the other hand, the Dry Highveld Grassland Spreads across the ordination space and is harder to classify accurately (Table 2).

The confidence with which modern biomes and bioregions are classified on the basis of pollen assemblages has implications for the application of the models to fossil pollen records and

predictions of past vegetation. Based on the evaluation metrics used here, accuracy, precision, F1 and Kappa statistics we are confident that our biome model correctly classifies three out of four biomes, namely: Savanna, Indian Ocean Coastal Belt, and Grassland (Table 1). For our bioregion model we are confident in correct classification of five out of seven bioregions, and particularly the Central Bushveld, Drakensberg Grassland, Eastern Kalahari Bushveld, Indian Ocean Coastal Belt and the Sub-Escarpment Grassland bioregions (Table 2).

4.2. Factors affecting the models' performance

There are several important factors that may affect models' performance. Accurate classification of a given vegetation class is reliant upon adequate representativeness of the class in the modern pollen record. This condition may be illustrated by the low performance on classification of Nama Karoo. Nama Karoo is represented by relatively small number of modern pollen samples (N = 18). We expect the classification performance to increase with more examples representative of Nama Karoo pollen assemblages. Likewise, the modern pollen samples from the Nama Karoo Biome used to train our classifier were collected from close proximity to the Grassland Biome. These surface samples may have received pollen input from neighboring grassland taxa by long distance dispersal which has been shown to exist (e.g. Scott and van Zinderen Bakker, 1985) and thus, may be more representative of a transitional zone rather than vegetation characteristic of the true Nama Karoo Biome.

Furthermore, the choice of the vegetation classification system used to assign vegetation labels to pollen data for training is likely to have an effect on the models' performance. Governed by a variety of global and local interacting factors, vegetation is not distributed evenly across the landscape. Although the classification system of Mucina and Rutherford (2006) used here is the most recent and detailed classification, it is restricted by international boundaries. Alternative vegetation classification systems could be used to broaden the geographic scope; however, other systems do not offer electronic maps resolved to comparable detail. For instance, Southern African vegetation is classified using a variety of requirements such as plant form and ecological characteristics (Low and Rebelo, 1996; Mucina and Rutherford, 2006; Rutherford and Westfall, 1986; Irish, 1994; Rutherford, 1997), species distribution (Linder et al., 2005, 2012; White, 1983) or disturbance factors like fire (Archibald et al., 2013). In addition, land use, agricultural practices and the introduction of exotic plants that have been continually changing the landscape (e.g. Acocks, 1988; Cronin et al., 2017; Hoffman and Todd, 2000; Hoffman & O'Connor, 1999; Hoffman and Cowling, 1990). Although best care was taken to sample surface pollen from relatively undisturbed sites these effects can never be completely eliminated. In consequence, attempts to classify vegetation, regardless of spatial scale, are imperfect thus introducing error in the models' predictions. Yet, despite possible imperfections of human-made classification systems and the lack of resolution in pollen identifications, the assemblages capture broad and regional vegetation communities enabling models for predictions of past vegetation states.

4.3. Important pollen taxa

We focus our discussion on pollen taxa that contribute significantly to accurate prediction of biomes and bioregions (Fig. 4, Table S.4). There is an overlap between the majority of the important pollen taxa identified by the two models. This overlap suggests that the taxa provide similar information to both models. This is expected as the classification of Southern African vegetation is hierarchical and bioregions are subsets of biomes. For instance, pollen of Amaranthaceae and *Tarchonanthus* is important to biome and bioregional predictions (Fig. 4 a and b). Pollen of Amaranthaceae family includes Chenopodiaceae, a pollen indicator of arid conditions (Scott, 1982b). Herbaceous Amaranthaceae are an important component of arid shurblands (Mucina et al., 2006c), while the distribution of these halophytic plants is associated with desert environments. In the modern pollen dataset Amaranthaceae is identified in 158 samples with high proportion characterizing dry vegetation types. Amaranthaceae is particularly prominent in Dry Highveld Grassland, Upper Karoo and Eastern Kalahari Bushveld (Fig. 3). Amaranthaceae is likely to be an important taxon in identifying relatively dry vegetation types. *Tarchonanthus* is a common woody shrub or small tree across Southern African savannas (Rutherford et al., 2006) and grasslands (Mucina et al., 2006b). *Tarchonanthus* pollen is identified in 71 modern pollen assemblages to a varying degree; it is an important pollen type of dry savanna vegetation but also present in other drier vegetation types (Fig. 3). As such, the taxon is likely to help in distinguishing between the Savanna and Grassland biomes. In the context of bioregions, relative proportions of *Tarchonanthus* pollen may be helpful in differentiating Eastern Kalahari Bushveld vs Central Bushveld as well as Sub-Escarpment vs Dry Highveld Grasslands.

*Anthospermum (*Rubiaceae) is an important taxon to the biome model (Fig. 4a). *Anthospermum,* a shrub closely resembling taxa typical of the Mediterranean-type vegetation of the Fynbos, is found in a variety of environments throughout South Africa including Indian Ocean Coastal Belt, Fynbos, as well as various types of grasslands (Mucina et al., 2006a), and savannas (Rutherford et al., 2006). In the modern pollen dataset *Anthospermum* is identified in 61 assemblages with highest abundances in Grassland biome and bioregions and to a lesser extent in Savanna biome and bioregions (Fig. 3). The importance of *Anthospermum* in the biome model is unexpected in view of its wide distribution over biomes.

Artemisia afra is an important taxon to the biome model (Fig. 4a). This low shrub is predominantly found in South African grasslands of the Eastern Free State but occurs also in savannas (Mucina et al., 2006a, b; Rutherford et al., 2006). Pollen of *A. afra*, identified in 41 modern assemblages, is characteristic of different types of grassland (Fig. 3). As such, *A. afra* is likely to be an important taxon for differentiating grassland bioregions. On the other hand, undifferentiated Combretaceae is important to classification of regional vegetation (Fig. 4b). Pollen of Combretaceae in the modern dataset most likely represents genera of deciduous trees such as *Combretum* and *Terminalia*. The family, often growing on sandy soils in association with *Burkea africana,* is a prominent component of both sub-humid and dry savannas and likely aids in their differentiation from other bioregions (Rutherford et al., 2006). In the modern pollen dataset, pollen of Combretaceae is characteristic of sub-humid savanna (Fig. 3). Thus, Combretaceae is likely to be an important taxon for distinguish savanna types such as the Central Bushveld and Eastern Kalahari Bushveld bioregions. B. africana is a deciduous frost-sensitive tree important in sub-humid savannas and typically growing on sandy soils (Rutherford et al., 2006). In modern assemblages, pollen of *B. africana* is almost exclusively restricted to sub-humid savanna (Fig. 3). Hence, the species most likely aids the model in discrimination between the Central Bushveld and the Eastern Kalahari Bushveld bioregions.

4.4. Biome shifts at Wonderkrater

Several factors may be at play to determine boundary shifts in savanna vegetation. The coexistence of trees and grass in contemporary savanna systems is a function of the combined influence and

feedback mechanisms between climate, resources and disturbance (House et al., 2003; Jeltsch et al., 2000; Sankaran et al., 2004; Scholes and Archer, 1997). For instance, intermediate precipitation promotes frequent fires through a positive feedback wherein the accumulation of biomass fuels the burning of grasses and maintains a savanna system (Daniau et al., 2012; van der Werf et al., 2004). Similarly, herbivory and browsing in particular, reduces shading by removing tree cover. Shading is an important determinant for savanna systems. Some savanna tree species, such as *Euclea divinorum*, require shade to establish their seedlings, while for other species, e.g. *Acacia tortilis*, *Acacia nilotica* or *Pappea capensis*, shaded conditions may be limiting or deadly (Holmes et al., 2008; Smith and Goodman, 1987; Smith and Shackleton, 1988). Ecological interactions on this level are generally not recognized in palynological studies but may be of importance in determining bioregions. Potentially, the machine learning methods can help us identify bioregion changes without having to consider these intricacies.

4.5. Late Pleistocene (c. 60–25 ka BP)

Previous studies suggest a series of oscillating events during the Late Pleistocene at Wonderkrater. The plant macrofossil records from Wonderkrater (Backwell et al., 2014; Bamford et al., 2016) as well as other pollen-based methods indicate a drying trend between 60–40 ka (Scott, 2016; Puech et al., 2017). Pollen-based quantitative reconstructions of temperature estimate a cooling of 2 °C at Wonderkrater throughout Marine Istope Stage 3 (MIS 3; Chevalier and Chase, 2015). Oscillations in moisture availability at Wonderkrater between 53–49 ka BP (Fig. 7 C) are thought to have led to changing conditions from relatively humid to drier (Scott, 2016). Predicted probabilities of the Grassland Biome support the notion of oscillating conditions between 60–40 ka BP. Furthermore, our bioregional model provides more detailed insights into regional vegetation and climatic conditions for this time period. Rather than dramatic shifts between savanna and grassland vegetation, our bioregional model suggests more nuanced vegetation changes in savanna composition from a relatively denser sub-humid savanna to a more open and drier savanna. A shift to the Eastern Kalahari Bushveld, a dry savanna type, is likely to be driven by reduced precipitation and slightly decreasing temperatures that may have led to a significantly higher incidence of frost (Mucina et al., 2006b).

Our models suggest that the period from c. 31 cal ka BP onwards marks the beginning of vegetation changes at Wonderkrater with savanna-grassland shifts (Fig. 4). Previous research suggests transitions from savanna to grassland (Scott, 1982a; Backwell et al., 2014). Pollen-based climate indices show a decrease in temperature and increase in moisture *c.* 26.5 cal ka BP (Scott, 2016). A recent pollen analysis from a new sequence at Wonderkrater also indicates cooler conditions with increased humidity before 30 cal ka BP (Puech et al., 2017). Our biome model reconstruction shows dominance of the Grassland Biome between *c.* 31-25 cal ka BP (Figs. 5 and 7 A). The Grassland Biome is associated with lower temperatures, increased precipitation, and higher incidence of frost (Mucina et al., 2006b). Our bioregional model predicts less frequent and abrupt shifts in regional vegetation with the proportion of the Drakensberg Grassland dominating at 26.5 cal ka BP (Figs. 6 and 7 B). A change on a vegetation continuum from sub-humid savanna to cool moist montane grassland would likely be indicative of significantly cooler temperatures leading to a significant increase in frost incidence (Mucina et al., 2006b). It is likely that frost was one of the important factors in maintaining Southern African temperate grasslands at a time of decreasing temperatures during terminal Pleistocene and leading up to the LGM (O'Conner & Bredenkamp, 1997).

4.6. Deglaciation (c. 20-15 cal ka BP)

Our biome assignment approach may be helpful in terms of reconciling disparate reconstructions and interpretations of the Wonderkrater pollen record during deglaciation at Wonderkrater. Pollen-based quantitative climate reconstructions suggest dry conditions during deglaciation (Chevalier and Chase, 2014; Truc et al., 2013). On the other hand, moisture curves derived from indicator pollen taxa have been interpreted as suggesting relatively humid condition (Fig. 7) (Scott, 1999a, 1999b; Scott et al., 2003, 2008, 2012). Our reconstructions of vegetation states at Wonderkrater suggest dominance of sub-humid savanna type (Figs. 6 and 7 B) indicative of warm and relatively moist conditions associated with contemporary climate. Considering declining probabilities of frost-dependent Eastern Kalahari Bushveld and Drakensberg Grassland, it is likely that temperature increased over this period.

4.7. Pleistocene-Holocene transition (c. 15–11.7 cal ka BP)

The terminal Pleistocene in central Southern Africa is somewhat enigmatic as a result of depositional hiatuses in regional isotope records (Brook et al., 2010; Holmgren et al., 2003) and poor representation in regional pollen records (Scott et al., 2012). In particular, abrupt climate changes in the northern hemisphere associated with the glacial re-advancement of the Younger Dryas (e.g. Alley, 2000; Grootes et al., 1993; Bond and Lotti, 1995) may not have occurred in Southern Africa (Abell and Plug, 2000; Scott et al., 1995). At Wonderkrater, we reconstruct significant changes in vegetation composition during the terminal Pleistocene as compared to Late Pleistocene (Fig. 7). Both of our biome and bioregional models indicate increased oscillations in the probabilities of all predicted biomes (Fig. 5) and bioregions (Fig. 6) between ~15.7 and 13.3 cal ka BP and again from ~12 cal ka BP into the early Holocene. Furthermore, the magnitude and frequency of these change is unparalleled as compared to our vegetation reconstructions for the Late Pleistocene. Specifically, the Pleistocene-Holocene transition is characterized by oscillations between Savanna and Grassland biomes and bioregions. The Wonderkrater pollen sequence for this time period records increased proportions of open savanna vegetation with arboreal Kalahari elements and dry grassland suggesting relatively dry conditions (Scott et al., 2003). Both the biome and bioregional reconstructions indicate higher probabilities for grassland development within the Younger Dryas period (Figs. 5–7).

However, quantitative reconstructions do not agree with respect to climatic conditions characterizing YD at Wonderkrater; while one study suggests cooling c. 12.3–10.5 cal ka BP (Truc et al., 2013), another estimates an increase in temperature, beginning c. 13.5, of up to 4 °C relative to the Glacial Maximum and highly variable precipitation until ~11 cal ka BP (Chevalier and Chase, 2015). The evidence for abrupt climate changes associated with Younger Dryas at Wonderkrater has been questioned as a result of low chronological resolution (Truc et al., 2013; Chevalier and Chase, 2015). Recently, the chronology of the Wonderkrater record has been optimized and recalibrated to improve the temporal resolution combining available boreholes (B1-4) (Scott, 2016). Here, our models use the updated chronology for vegetation reconstructions. Given the abrupt and highly variable nature along with timing of our reconstructed trends in probability of vegetation types at Wonderkrater, we suggest they may be driven by abrupt climatic changes associated with YD.

4.8. The Holocene (c. 11.7 cal ka BP to present)

Reconstructed trends for Late Pleistocene vegetation continue well into the Holocene with more frequent and higher amplitude

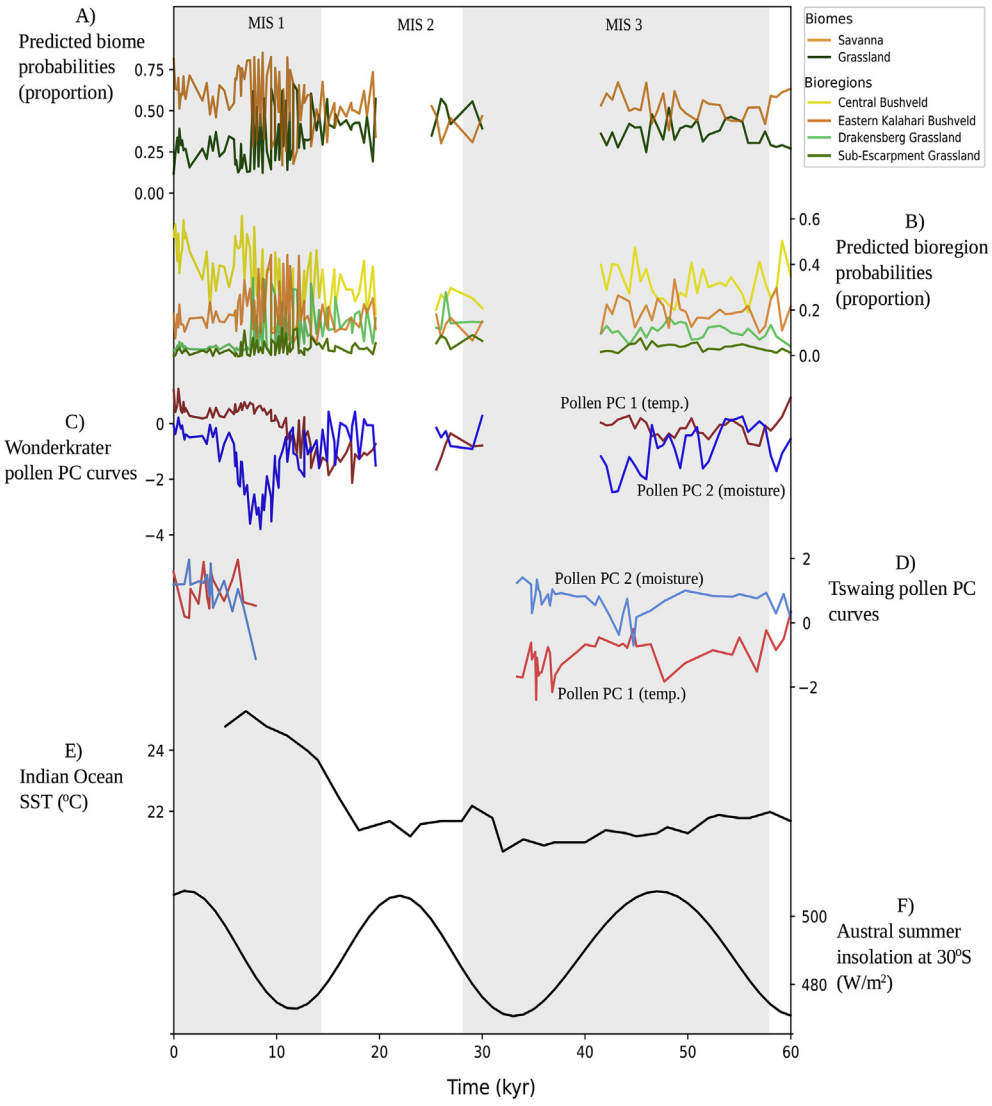


Fig. 7. Comparison of reconstructed biome (A) and bioregions (B) at Wonderkrater with regional proxies: C) Wonderkrater pollen principal components curves (Scott, 2016), D) Tswaing pollen principal components curves (Scott, 2016), E) Indian Ocean sea surface temperatures (Brad & Rickaby, 2009), F) December 30°S precession cycle (Berger and Loutre, 1991).

changes in vegetation communities (Figs. 5 and 6). Oscillating trends in the probability of biomes and bioregions suggest more variable conditions than earlier in the record. Between 9.5 and 6 cal ka BP our reconstruction indicates Eastern Kalahari Bushveld alternating with Drakensberg Grassland with an eventual transition to Central Bushveld (Mucina et al., 2006b). Increases in the probability of the Kalahari savanna type as well as montane grassland (Figs. 6 and 7 B) indicate variable conditions likely characterized by fluctuating temperatures and inconsistent precipitation (Mucina et al., 2006b). Our reconstructions are consistent with pollen spectra that show an increase in the relative presence of dry woodland vegetation between 9.5 and 6 cal ka BP along with a decreasing trend in a pollen-based moisture index (Scott et al., 2003). However, increased precipitation is reconstructed starting ~11.5 and continuing to mid-Holocene in other studies (Truc et al., 2013; Chevalier and Chase, 2014).

From the mid-Holocene onwards, vegetation trends begin to stabilize as the sub-humid Central Bushveld savanna with a grassland component establishes at Wonderkrater (Fig. 7 A and B). Decreasing probabilities of Grassland and dry Eastern Kalahari Bushveld (Fig. 7 panels A and B), both characterized by colder temperatures and higher incidence of frost (Mucina et al., 2006b), suggest more mild local conditions similar to the contemporary regional climate (Rutherford et al., 2006). Furthermore, our regional vegetation model indicates the occurrence of a dry event during the late Holocene represented by an increase in probability of the dry savanna at 3.5 cal ka BP. Pollen analysis shows increased proportions of fern spores at Wonderkrater beginning ~6.5 cal ka BP along with increasing moisture index peaking at 5.5 cal ka BP interpreted as suggesting more humid local conditions (Scott et al., 2003). Quantitative climate reconstructions support increased humidity with maximum precipitation occurring between 7 and 3 cal ka BP (Truc et al., 2013; Chevalier and Chase, 2015). Reconstructed estimates of past temperatures show a rising trend that peaks ~7.5 cal ka BP followed by a decline in temperatures after 6.5 cal ka BP (Truc et al., 2013; Chevalier and Chase, 2015). A dry event ~3 cal ka BP is inferred from the pollen record showing increased proportion of Asteraceae pollen (Scott, 1982a), and supported by a decreasing trend in reconstructed temperature estimates (Truc et al., 2013; Chevalier and Chase, 2015).

5. Conclusions

We apply new machine learning-based models to modern pollen assemblages and a fossil pollen sequence for reconstructions of past vegetation states in southern African. Our data-driven approach is a promising method for more objective and probabilistic reconstructions of past vegetation histories from fossil pollen records. As such, our method may be helpful in reconciling disparate reconstructions obtained using other methods. For additional insights into past environments, the machine learning approach may be used in conjunction with other methods such as traditional pollen analysis or pollen-based quantitative climate reconstructions.

We acknowledge that the performance of our models is dependent upon large modern pollen dataset and adequate representation of different vegetation types. Vegetation types that are well represented in the modern pollen data, such as savanna, allow accurate reconstructions of past vegetation. Thus, we stress the importance of high quality empirical data, both modern and fossil, upon which modeling studies are conditional. This is particularly important in paleosciences for which reconstructions of past conditions, whether qualitative or quantitative, are only as accurate as the empirical data available.

6. Data availability

Datasets and R code related to this article can be found at https://github.com/Betelgesse/SAModernPollen hosted on GitHub.

Acknowledgments

This research was supported by funding to M.K. Sobol from the Ontario Graduate Scholarship (OGS) and the Queen Elizabeth II Graduate Scholarship in Science and Technology (QEII-GSST); the National Research Foundation (NRF, South Africa, grant no. 85903) grant to L. Scott; and grants to S.A. Finkelstein from the Natural Sciences and Engineering Research Council of Canada. We would like to thank Dr Graciela Gil Romera, Dr Eugene Marais, and Dr Graham Avery for contributing modern pollen data. J. S. Rudy for guidance during analyses and interpretations, and A.K. Phillips for useful discussions and comments on the manuscript. Special thanks to Liora Kolska Horwitz and Michael Chazan for arranging a research visit to Bloemfontein for M.K. Sobol that provided the opportunity to work hands-on with modern pollen reference collection. Any opinion, finding, and conclusion or recommendation expressed in this material is that of the authors and the NRF does not accept any liability in this regard.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.quascirev.2019.03.027.

References

Acocks, J.P.H., 1988. Veld Types of South Africa: Memoirs of the Botanical Survey of South Africa, No. 57. Botanical Research Institute, Department of Agriculture and Water Supply, South Africa.

Abell, P.I., Plug, I., 2000. The Pleistocene/Holocene transition in South Africa: evidence for the Younger Dryas event. Glob. Planet. Chang. 26 (1–3), 173–179. https://doi.org/10.1016/S0921-8181(00)00042-4.

Alley, R.B., 2000. The Younger Dryas cold interval as viewed from central Greenland. Quat. Sci. Rev. 19 (1–5), 213–226. https://doi.org/10.1016/S0277-3791(99)00062-1.

Archer, S., Schimel, D.S., Holland, E.A., 1995. Mechanisms of shrubland expansion: land use, climate or CO2. Clim. Change (2), 91–99. Retrieved from. https://link.springer.com/content/pdf/10.1007%2FBF01091640.pdf.

Archibald, S., Lehmann, C.E., Gómez-Dans, J.L., Bradstock, R.A., 2013. Defining pyromes and global syndromes of fire regimes. Proc. Natl. Acad. Sci. Unit. States Am. 110 (16), 6442–6447. https://doi.org/10.1073/pnas.121.

Backwell, L.R., McCarthy, T.S., Wadley, L., Henderson, Z., Steininger, C.M., deKlerk, Bonita, et al., 2014. Multiproxy record of late Quaternary climate change and Middle Stone Age human occupation at Wonderkrater, South Africa. Quat. Sci. Rev. 99, 42–59. https://doi.org/10.1016/j.quascirev.2014.06.017.

Bamford, M.K., Neumann, F.H., Scott, L., 2016. Pollen, Charcoal and Plant Macrofossil Evidence of Neogene and Quaternary Environments in Southern Africa. Quaternary Environmental Change in Southern Africa. Cambridge University Press, Cambridge, pp. 306–323.

Berger, A., Loutre, M.F., 1991. Insolation values for the climate of the last 10 million years. Quat. Sci. Rev. 10 (4), 297–317. https://doi.org/10.1016/0277-3791(91)90033-Q.

Blaauw, M., Christeny, J.A., 2011. Flexible paleoclimate age-depth models using an autoregressive gamma process. Bayesian Analysis 6 (3), 457–474. https://doi.org/10.1214/11-BA618.

Bond, G.C., Lotti, R., 1995. Iceberg discharges into the North Atlantic on millennial time scales during the last glaciation. Science 267 (5200), 1005–1010. https://doi.org/10.1126/science.267.5200.1005.

Brad, Edouard, Rickaby, Rosalind EM, 2009. Migration of the subtropical front as a modulator of glacial climate. Nature 460 (7253), 380.

Brook, G.A., Scott, L., Railsback, B., Goddard, E.A., 2010. A 35 ka pollen and isotope record of environmental change along the southern margin of the Kalahari from a stalagmite in Wonderwerk Cave, South Africa. J. Arid Environ. 74 (5), 870–884. https://doi.org/10.1016/j.jaridenv.2009.11.006.

Carrion, J.S., Scott, L., Vogel, J.C., 1999. Twentieth century changes in montane vegetation in the eastern Free State, South Africa, derived from palynology of hyrax dung middens. J. Quat. Sci. 14 (1), 1–16, 1<1::AID-JQS412>3.0.CO;2-Y. https://doi.org/10.1002/(SICI)1099-1417(199902)14.

Chevalier, M., Chase, B.M., 2015. Southeast African records reveal a coherent shift from high- to low-latitude forcing mechanisms along the east African margin across last glacial-interglacial transition. Quat. Sci. Rev. 125, 117–130. https://doi.org/10.1016/j.quascirev.2015.07.009.

Chevalier, M., Chase, B.M., 2016. Determining the drivers of long-term aridity variability: a southern African case study. J. Quat. Sci. 31 (2), 143–151. https://doi.org/10.1002/jqs.2850.

Chevalier, M., Cheddadi, R., Chase, B.M., 2014. CREST (Climate REconstruction SofTware): a probability density function (PDF)-based quantitative climate reconstruction method. Clim. Past 10 (6), 2081–2098. https://doi.org/10.5194/cp-10-2081-2014.

Cooremans, B., 1989. Pollen production in central southern Africa. Pollen Spores 36, 61–78.

Cronin, K., Kaplan, H., Gaertner, M., Irlich, U., Hoffman, M.T., 2017. Aliens in the nursery: assessing the attitudes of nursery managers to invasive species regulations. Biol. Invasions 19 (3), 925–937. https://doi.org/10.1007/s10530-016-1363-3.

Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. Ecology 88 (11), 2783–2792. https://doi.org/10.1890/07-0539.1.

Daniau, A.L., Bartlein, P.J., Harrison, S.P., Prentice, I.C., Brewer, S., Friedlingstein, P., Harrison-Prentice, T.I., Inoue, J., Izumi, K., Marlon, J.R., Mooney, S., 2012. Predictability of biomass burning in response to climate changes. Glob. Biogeochem. Cycles 26 (4), 2014–2015. https://doi.org/10.1029/2011GB004249.

Fægri, K., Iversen, J., 1986. Textbook of Pollen Analysis, fourth ed. Wiley, Chichester.

Grootes, P.M., Stuiver, M., White, J.W.C., Johnsen, S., Jouzel, J., 1993. Comparison of oxygen isotope records from the GISP2 and GRIP Greenland ice cores. Nature 366 (6455), 552. https://doi.org/10.1038/366552a0.

Higgins, S.I., Scheiter, S., 2012. Atmospheric CO2 forces abrupt vegetation shifts locally, but not globally. Nature 488 (7410), 209. https://doi.org/10.1038/nature11238.

Hill, M.O., Gauch, H.G., 1980. Detrended correspondence analysis: an improved ordination technique. Vegetatio 42 (1–3), 47–58. https://doi.org/10.1007/BF00048870.

Hoffman, M.T., Todd, S.W., 2000. A national review of land degradation in South Africa: the influence of biophysical and socio-economic factors. J. South. Afr. Stud. 26 (4), 743–758.

Hoffman, M.T., O'Connor, T.G., 1999. Vegetation change over 40 years in the Weenen/Muden area, KwaZulu-Natal: evidence from photo-panoramas. Afr. J. Range Forage Sci. 16 (2&3), 71–88.

Hoffman, M.T., Cowling, R.M., 1990. Vegetation change in the semi-arid, eastern Karoo over the last two hundred years: an expanding Karoo - fact or fiction? South Afr. J. Sci. 86, 286–294.

Holmes, P.J., Bateman, M.D., Thomas, D.S.G., Telfer, M.W., Barker, C.H., Lawson, M.P., 2008. A Holocene-late Pleistocene aeolian record from lunette dunes of the western Free State panfield, South Africa. Holocene 18 (8), 1193–1205. https://doi.org/10.1177/0959683608095577.

Holmgren, K., Lee-Thorp, J.A., Cooper, G.R., Lundblad, K., Partridge, T.C., Scott, L., Sithaldeen, R., Talma, A.S., Tyson, P.D., 2003. Persistent millennial-scale climatic variability over the past 25,000 years in Southern Africa. Quat. Sci. Rev. 22 (21–22), 2311–2326. https://doi.org/10.1016/S0277-3791(03)00204-X.

House, J.I., Archer, S., Breshears, D.D., Scholes, R.J., NCEAS Tree–Grass Interactions Participants, 2003. Conundrums in mixed woody-herbaceous plant systems. J. Biogeogr. 30 (11), 1763–1777. https://doi.org/10.1046/j.1365-2699.2003.00873.x.

Irish, J., 1994. The biomes of Namibia as determined by objective categorisation.

Navors. Nas. Mus. (Bloemfontein): Res. Natl. Mus. (Bloemfontein) 10 (13), 549–592.
Jeltsch, F., Weber, G.E., Grimm, V., 2000. Ecological buffering mechanisms in savannas: a unifying theory of long-term tree-grass coexistence. Plant Ecol. 150 (1–2), 161–171. https://doi.org/10.1023/A:102659080668.
Jolly, D., Prentice, I.C., Bonnefille, R., Ballouche, A., Bengo, M., Brenac, P., Buchet, G., Burney, D., Cazet, J.P., Cheddadi, R., Edorh, T., 1998. Biome reconstruction from pollen and plant macrofossil data for Africa and the Arabian peninsula at 0 and 6000 years. J. Biogeogr. 25 (6), 1007–1027. https://doi.org/10.1046/j.1365-2699.1998.00238.x.
Lebamba, J., Ngomanda, A., Vincens, A., Jolly, D., Favier, C., Elenga, H., Bentaleb, I., 2009. A reconstruction of Atlantic Central African biomes and forest succession stages derived from modern pollen data and plant functional types. Clim. Past 5 (3), 403–429. https://doi.org/10.5194/cp-5-403-2009.
Lézine, A., Watrin, J., Vincens, A., Hély, C., 2009. Are modern pollen data representative of west African vegetation ? Rev. Palaeobot. Palynol. 156 (3), 256–276. https://doi.org/10.1016/j.revpalbo.2009.02.001.
Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R. News 2 (December), 18–22. https://doi.org/10.1177/154405910408300516.
Linder, H.P., de Klerk, H.M., Born, J., Burgess, N.D., Fjeldså, J., Rahbek, C., 2012. The partitioning of Africa: statistically defined biogeographical regions in sub-Saharan Africa. J. Biogeogr. 39 (7), 1189–1205. https://doi.org/10.1111/j.1365-2699.2012.02728.x.
Linder, H.P., Lovett, J., Mutke, J.M., Barthlott, W., Jürgens, N., Rebelo, T., Küper, W., 2005. A numerical re-evaluation of the sub-Saharan phytochoria of mainland Africa. Biol. Skr. 55, 229–252.
Low, A.B., Rebelo, A.G., 1996. Vegetation of South Africa, Lesotho and Swaziland. Department of Environmental Affairs and Tourism, Pretoria, South Africa.
Mucina, L., Hoare, D.B., Lötter, M.C., du Preez, P.J., Rutherford, M.C., Scott-Shaw, C.R., Bredenkamp, G.J., Powrie, L.W., Scott, L., Cilliers, S.S., Bezuidenhout, H., Mostert, T.H., Camp, K.G.T., Siebert, S.J., Winter, P.J.D., Burrows, J.E., Dobson, L., Ward, R.A., Stalmans, M., Oliver, E.G.H., Siebert, F., Kobisi, K., Kose, L., 2006a. Grassland Biome. In: & Mucina, L., Rutherford, M. (Eds.), Vegetation of South Africa, Lesotho & Swaziland. Strelitzia, vol. 19, pp. 349–437.
Mucina, L., Rutherford, M.C. (Eds.), 2006. The Vegetation of South Africa, Lesotho and Swaziland. Strelitzia, vol. 19.
Mucina, L., Rutherford, M.C., Powrie, L.W., 2006b. Biomes and bioregions of Southern Africa. In: Mucina, L., Rutherford, M.C. (Eds.), The Vegetation of South Africa, Lesotho and Swaziland. Strelitzia, vol. 19, pp. 31–51.
Mucina, L., Rutherford, M.C., Palmer, A.R., Milton, S.J., Scott, L., Lloyd, J.W., Van der Merwe, B., Hoare, D.B., Bezuidenhout, H., Vlok, J.H.J., Euston-Brown, D.I.W., 2006c. Nama-karoo Biome. In: Mucina, L., Rutherford, M.C. (Eds.), The Vegetation of South Africa, Lesotho and Swaziland. Strelitzia, vol. 19, pp. 324–347.
Mucina, L., Scott-Shaw, C.R., Rutherford, M.C., Camp, K.G.T., Matthews, W.S., Powrie, L.W., Hoare, D.B., 2006d. Indian Ocean Coastal Belt. In: Mucina, L., Rutherford, M.C. (Eds.), The Vegetation of South Africa, Lesotho and Swaziland. Strelitzia, vol. 19, pp. 569–583.
O'Connor, T.G., Bredenkamp, G.J., 1997. Grassland. Vegetation of Southern Africa. Cambridge University Press, Cambridge, pp. 215–257.
O'Connor, T.G., Puttick, J.R., Hoffman, M.T., 2014. Bush encroachment in southern Africa: changes and causes. Afr. J. Range Forage Sci. 31 (2), 67–88. https://doi.org/10.2989/10220119.2014.939996.
Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., O'hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H., 2017. Vegan: community ecology package. Vegan: community ecology package. https://doi.org/ISBN 0-387-95457-0.
Palmer, A.R., Hoffman, M.T., 1997. Nama Karoo. In: Cowling, R.M., Richardson, D.M., Pierce, S.M. (Eds.), Vegetation of Southern Africa. Cambridge University Press, Cambridge, pp. 167–186.
Prentice, I.C., Cramer, W., Harrison, S.P., Leemans, R., Monserud, R.A., Solomon, A.M., 1992. A global biome model based on plant physiology and dominance, soil properties and climate. J. Biogeogr. 19 (2), 117–134. https://doi.org/10.2307/2845499.
Prentice, I.C., Guiot, J., Huntley, B., Jolly, D., Cheddadi, R., 1996. Reconstructing biomes from palaeoecological data: a general method and its application to European pollen data at 0 and 6 ka. Clim. Dyn. 12, 185–194. https://doi.org/10.1007/s003820050102.
Puech, E., Urrego, D.H., Sánchez Goñi, M.F., Backwell, L., d'Erricoc, F., 2017. Vegetation and environmental changes at the middle stone age site of wonderkrater, Limpopo, South Africa. Quat. Res. 88 (2), 313–326. https://doi.org/10.1017/qua.2017.42.
Rutherford, M.C., 1982. Annual production fraction of aboveground biomass in relation to plant shrubbiness in savanna. Bothalia 14 (1), 139–142.
Rutherford, M.C., 1997. Categorization of biomes. In: Cowling, R.M., Richardson, D.M., Pierce, S.M. (Eds.), Vegetation of Southern Africa. Cambridge University Press, Cambridge, UK, pp. 91–98.
Rutherford, M.C., Mucina, L., Lötter, C., Bredenkamp, G.J., Smit, J.H.L., Scott-Shaw, C.R., Hoare, D.B., Goodman, P.S., Bezuidenhout, H., Scott, L., Ellis, F., Powrie, L.W., Siebert, F., Mostert, T.H., Henning, B.J., Venter, C.E., Camp, K.G.T., Siebert, S.J., Matthews, W.S., Burrows, J.E., Dobson, L., van Rooyen, N., Schmidt, E., Winter, P.J.P., du Preez, P.J., Ward, R.A., Williamson, S.W., Hurter, P.J.H., 2006. Savanna Biome. In: Mucina, L., Rutherford, M.C. (Eds.), Vegetation of South Africa, Lesotho & Swaziland. SANBI, Pretoria, pp. 439–539. https://doi.org/10.1007/s.
Rutherford, M.C., Westfall, R.H., 1986. Biomes of southern Africa - an objective categorization. Mem. Bot. Surv. S. Afr. 54, 1–98.
Sala, O.E., 2000. Global biodiversity scenarios for the year 2100. Science 287 (5459), 1770–1774. https://doi.org/10.1126/science.287.5459.1770.
Sankaran, M., Hanan, N.P., Scholes, R.J., Ratnam, J., Augustine, D.J., Cade, B.S., Bicini, G., Bronn, A., Worden, J., Ekaya, W., Feral, C.J., Banyikwa, F., Prins, H.H.T., Ringrose, S., Gignoux, J., Diouf, A., Higgins, S.I., February, E.C., Sea, W., Ludwig, F., Hiernaux, P., Hrabar, H.,, Frost, P.G.H., Tews, J., Coughenour, M.B., Zambatis, N., Ratnam, J., Caylor, K.K., Ardo, J., Scholes, R.J., Roux, X., Metzger, K.L., 2005. Determinants of woody cover in African savannas. Nature 438 (7069), 846–849. https://doi.org/10.1038/nature04070.
Sankaran, M., Ratnam, J., Hanan, N.P., 2004. Tree-grass coexistence in savannas revisited - insights from an examination of assumptions and mechanisms invoked in existing models. Ecol. Lett. 7 (6), 480–490. https://doi.org/10.1111/j.1461-0248.2004.00596.x.
Scholes, R.J., Archer, S.R., 1997. Tree-grass interactions in savannas. Annu. Rev. Ecol. Systemat. 28 (1), 517–544. https://doi.org/10.1146/annurev.ecolsys.28.1.517.
Scholes, R.J., Walker, B.H., 1993. An African Savanna: Synthesis of the Nylsvley Study. Cambridge University Press, Cambridge.
Scott, L., 1982a. A late quaternary pollen record from the Transvaal bushveld, South Africa. Quat. Res. 17 (3), 339–370. https://doi.org/10.1016/0033-5894(82)90028-X.
Scott, L., 1982b. Late quaternary fossil pollen grains from the Transvaal, South Africa. Rev. Palaeobot. Palynol. 36 (3–4), 241–278. https://doi.org/10.1016/0034-6667(82)90022-7.
Scott, L., 1989. Pollen analysis and palaeoenvironmental interpretation of Late Quaternary sediment exposures in the eastern Orange Free State, South Africa. South Afr. J. Bot. 55 (1), 107–116.
Scott, L., 1999a. Palynological analysis of the Pretoria Saltpan (Tswaing crater) sediments and vegetation history of the bushveld savanna biome, South Africa. In: Partridge, T. (Ed.), In Tswaing – Investigations into the Origin, Age and Palaeoenvironments of the Pretoria Saltpan. Council of Geoscience, Pretoria, pp. 143–166.
Scott, L., 1999b. Vegetation history and climate in the Savanna Biome South Africa since 190,000 ka: a comparison of pollen data from the Tswaing Crater (the Pretoria Saltpan) and Wonderkrater. Quat. Int. 57 (58), 215–223. https://doi.org/10.1016/S1040-6182(98)00062-7.
Scott, L., 2016. Fluctuations of vegetation and climate over the last 75 000 years in the Savanna Biome, South Africa: Tswaing Crater and Wonderkrater pollen sequences reviewed. Quat. Sci. Rev. 145, 117–133. https://doi.org/10.1016/j.quascirev.2016.05.035.
Scott, L., Cooremans, B., 1992. Pollen in recent Procavia (hyrax), Petromus (dassie rat) and bird dung in South Africa. J. Biogeogr. 19 (2), 205. https://doi.org/10.2307/2845506.
Scott, L., Cooremans, B., Maud, R.R., 1992. Preliminary palynological evaluation of the port durnford formation at Port Durnford, Natal coast, South Africa. South Afr. J. Sci. 88, 470–474.
Scott, L., Holmgren, K., Partridge, T.C., 2008. Reconciliation of vegetation and climatic interpretations of pollen profiles and other regional records from the last 60 thousand years in the Savanna Biome of Southern Africa. Palaeogeogr. Palaeoclimatol. Palaeoecol. 257 (1–2), 198–206. https://doi.org/10.1016/j.palaeo.2007.10.018.
Scott, L., Holmgren, K., Talma, A.S., Woodborne, S., Vogel, J.C., 2003. Age interpretation of the Wonderkrater spring sediments and vegetation change in the Savanna Biome, Limpopo province, South Africa. South Afr. J. Sci. 99 (9–10), 484–488.
Scott, L., Neumann, F.H., Brook, G.A., Bousman, C.B., Norström, E., Metwally, A.A., 2012. Terrestrial fossil-pollen evidence of climate change during the last 26 thousand years in Southern Africa. Quat. Sci. Rev. 32, 100–118. https://doi.org/10.1016/j.quascirev.2011.11.010.
Scott, L., Steenkamp, M., Beaumont, P.B., 1995. Palaeoenvironmental conditions in South Africa at the Pleistocene-Holocene transition. Quat. Sci. Rev. 14 (9), 937–947. https://doi.org/10.1016/0277-3791(95)00072-0.
Scott, L., Thackeray, J.F., 1987. Multivariate analysis of late Pleistocene and holocene pollen spectra from wonerkrater, transvaal, South Africa. South Afr. J. Sci. 83 (2), 93–97.
Scott, L., van Zinderen Barker, E.M., 1985. Exotic pollen and long-distance wind dispersal at a sub-Antarctic Island. Grana 24 (1), 45–54. https://doi.org/10.1080/00173138509427422.
Simpson, G.L., Birks, H.J.B., 2012. Statistical learning in palaeolimnology. In: Tracking Environmental Change Using Lake Sediments. Springer, Dordrecht, pp. 249–327.
Skowno, A.L., Thompson, M.W., Hiestermann, J., Ripley, B., West, A.G., Bond, W.J., 2017. Woodland expansion in South African grassy biomes based on satellite observations (1990–2013): general patterns and potential drivers. Glob. Chang. Biol. 23 (6), 2358–2369. https://doi.org/10.1111/gcb.13529.
Smith, T.M., Goodman, P.S., 1987. Successional dynamics in an *Acacia nilotica-Euclea divinorum* savannah in southern Africa. J. Ecol. 603–610. https://doi.org/10.2307/2260192.
Smith, T.M., Shackleton, S.E., 1988. The effects of shading on the establishment and growth of Acacia tortilis seedlings. South Afr. J. Bot. 54 (4), 375–379. https://doi.org/10.1016/S0254-6299(16)31305-9.
Sobol, M.K., Finkelstein, S.A., 2018. Predictive pollen-based biome modeling using machine learning. PLoS One 13 (8) e0202214. https://doi.org/10.1371/journal.pone.0202214.
Sowunmi, M.A., 1973. Pollen grains of Nigerian plants: I. Woody species. Grana 13,

145–186. https://doi.org/10.1080/00173137309429891.
Sowunmi, M.A., 1995. Pollen of Nigerian plants pollen of Nigerian plants: II. Woody species. Grana 34, 120–141. https://doi.org/10.1080/00173139509430002.
Stevens, N., Lehmann, C.E., Murphy, B.P., Durigan, G., 2017. Savanna woody encroachment is widespread across three continents. Glob. Chang. Biol. 23 (1), 235–244. https://doi.org/10.1111/gcb.13409.
Staver, A.C., Archibald, S., Levin, S., 2011. Tree cover in sub-Saharan Africa: rainfall and fire constrain forest and savanna as alternative stable states. Ecology 92 (5), 1063–1072. https://doi.org/10.1890/i0012-9658-92-5-1063.
Stockmarr, J., 1971. Tablets with spores used in absolute pollen analysis. Pollen Spores 13, 615–621.
Truc, L., Chevalier, M., Favier, C., Cheddadi, R., Meadows, M.E., Scott, L., Chase, B.M., 2013. Quantification of climate change for the last 20,000years from Wonderkrater, South Africa: implications for the long-term dynamics of the intertropical convergence zone. Palaeogeogr. Palaeoclimatol. Palaeoecol. 386, 575–587. https://doi.org/10.1016/j.palaeo.2013.06.024.
van der Werf, G.R., Randerson, J.T., Collatz, G.J., Giglio, L., Kasibhatla, P.S., Arellano Jr., A.F., Olsen, S.C., Kasischke, E.S., 2004. Continental-scale partitioning of fire emissions during 1997 to 2001 El Nino/La Nina period. Science 303 (5654), 73–76. https://doi.org/10.1126/science.1090753.
van Geel, B., 1978. A palaeoecological study of Holocene peat bog sections in Germany and The Netherlands, based on the analysis of pollen, spores and macro- and microscopic remains of fungi, algae, cormophytes and animals. Rev. Palaeobot. Palynol. 25 (1), 1–120. https://doi.org/10.1016/0034-6667(78)90040-4.
van Geel, B., Aptroot, A., 2006. Fossil ascomycetes in Quaternary deposits. Nova Hedwigia 82 (3–4), 313–329. https://doi.org/10.1127/0029-5035/2006/0082-0313.
van Zinderen Bakker, E.M., 1953. South African Pollen Grains and Spores I. AA Balkema, Cape Town, p. 88.
van Zinderen Bakker, E.M., 1956. South African Pollen Grains and Spores II. AA Balkema, Cape Town, p. 132.
van Zinderen Bakker, E.M., Coetzee, J.,A., 1959. South African Pollen Grains and Spores III. AA Balkema, Cape Town, p. 200.
van Zinderen Bakker, E.M., Welman, M., Kuhn, L., 1970. South African Pollen Grains and Spores IV. AA Balkema, Cape Town, p. 110.
Vincens, A., Bremond, L., Brewer, S., Buchet, G., Dussouillez, P., 2006. Modern pollen-based biome reconstructions in East Africa expanded to southern Tanzania. Rev. Palaeobot. Palynol. 140 (3–4), 187–212. https://doi.org/10.1016/j.revpalbo.2006.04.003.
Verlhac, L., Izumi, K., Lézine, A., Lemonnier, K., Buchet, G., Achoundong, G., Tchiengué, B., 2018. Altitudinal distribution of pollen , plants and biomes in the Cameroon highlands. Rev. Palaeobot. Palynol. 259, 21–28. https://doi.org/10.1016/j.revpalbo.2018.09.011.
Ward, D., 2005. Do we understand the cases of bush encroachment in African savannas? Afr. J. Range Forage Sci. 22 (2), 101–105. https://doi.org/10.2989/10220110509485867.
White, F., 1983. The Vegetation of Africa: a Descriptive Memoir to Accompany the UNESCO/AETFAT/UNSO Vegetation Map of Africa, twentieth ed. Natural Resources Research, Paris, France.
Williams, J.W., Shuman, B.N., Webb III, T., Bartlein, P.J., Leduc, P.L., 2004. Late-Quaternary vegetation dynamics in North America: scaling from taxa to biomes. Ecol. Monogr. 74 (2), 309–334.